MONTHLY NOTICES

OF THE

ROYAL ASTRONOMICAL SOCIETY

Vol. 115 No. 6 1955

Published and Sold by the
ROYAL ASTRONOMICAL SOCIETY
BURLINGTON HOUSE
LONDON, W.1

Price Thirteen Shillings and Sixpence

NOTICE TO AUTHORS

1. *Communications.*—Papers must be communicated to the Society by a Fellow. They should be accompanied by a summary at the *beginning* of the paper conveying briefly the content of the paper, and drawing attention to important new information and to the main conclusions. The summary should be intelligible in itself, without reference to the paper, to a reader with some knowledge of the subject; it should not normally exceed 200 words in length. **Authors are requested to submit MSS. in duplicate. These should be typed using double spacing and leaving a margin of not less than one inch on the left-hand side. Corrections to the MSS. should be made in the text and not in the margin.** Unless a paper reaches the Secretaries more than seven days before a Council meeting it will not normally be considered at that meeting. By Council decision, MSS. of accepted papers are retained by the Society for one year after publication; unless their return is then requested by the author, they are destroyed.

2. *Presentation.*—Authors are allowed considerable latitude, but they are requested to follow the general style and arrangement of *Monthly Notices*. References to literature should be given in the standard form, including a date, for printing either as footnotes or in a numbered list at the end of the paper. Each reference should give the **name and initials** of the author cited, irrespective of the occurrence of the name in the text (some latitude being permissible, however, in the case of an author referring to his own work). The following examples indicate the style of reference appropriate for a paper and a book, respectively :—

A. Corlin, *Zs. f. Astrophys.*, **15**, 239, 1938.
H. Jeffreys, *Theory of Probability*, 2nd edn., section 5.45, p. 258, Oxford, 1948.

3. *Notation.*—For technical astronomical terms, authors should conform closely to the recommendations of Commission 3 of the International Astronomical Union (*Trans. I.A.U.*, Vol. VI, p. 345, 1938). Council has decided to adopt the I.A.U. 3-letter abbreviations for constellations where contraction is desirable (Vol. IV, p. 221, 1932). In general matters, authors should follow the recommendations in *Symbols, Signs and Abbreviations* (London : Royal Society, 1951) except where these conflict with I.A.U. practice.

4. *Diagrams.*—**These should be designed to appear upright on the page, drawn about twice the size required in print and prepared for direct photographic reproduction except for the lettering, which should be inserted in pencil. Legends should be given in the manuscript indicating where in the text the figure should appear.** Blocks are retained by the Society for 10 years; unless the author requires them before the end of this period they are then destroyed.

5. *Tables.*—**These should be arranged so that they can be printed upright on the page.**

6. *Proofs.*—Costs of alteration exceeding 5 per cent of composition must be borne by the author. Fellows are warned that such costs have risen sharply in recent years, and it is in their own and the Society's interests to seek the maximum conciseness and simplification of symbols and equations consistent with clarity.

7. *Revised Manuscripts.*—When papers are submitted in revised form it is especially requested that they be accompanied by the original MSS.

Reading of Papers at Meetings

8. When submitting papers authors are requested to indicate whether they will be willing and able to read the paper at the next or some subsequent meeting, and approximately how long they would like to be allotted for speaking.

9. Postcards giving the programme of each meeting are issued some days before the meeting concerned. Fellows wishing to receive such cards whether for Ordinary Meetings or for the Geophysical Discussions or both should notify the Assistant Secretary.

MONTHLY NOTICES

OF THE

ROYAL ASTRONOMICAL SOCIETY

Vol. 115 No. 6

ADDITIONAL MEETING OF 1955 SEPTEMBER 8

in the Chemistry Lecture Theatre of
the University of Glasgow

Professor Sir Harold Jeffreys, President, in the Chair

The President announced that this Additional Meeting was the seventh to take place outside the rooms of the Society in London. He expressed the thanks of the Society to Professor W. M. Smart, at whose invitation the meeting was being held at Glasgow, and to Professor Robertson for the use of the Chemistry Lecture Theatre. The thanks of the Society were also extended to the Warden of Queen Margaret Hall, where accommodation had been provided for Fellows participating in the various activities arranged in connection with the visit, and to Dr M. W. Ovenden, who had undertaken much of the work involved in the local organization of the meetings and associated activities.

Two hundred and twenty-one presents were announced as having been received since the last meeting, including:—

F. G. H. Struve, *Beobachtungen des Halleyschen Cometen* (presented by Miss D. Davies);

E. H. Linfoot, *Recent Advances in Optics* (presented by the author);

J. Picard, *Abhandlunger vom Wasserwagen*, *Ouvrages de mathématique*, *Degré du méridien entre Paris et Amiens*, *Traité du nivellement*; G. Mouton, *Observationes diametrorum Solis et Lunae apparentium*; Académie Royale des Sciences, Paris, *Recueil de plusieurs traites de mathematique*, *Recueil d'observations faites en plusieurs voyages*, *Memoirs for a natural history of animals*; P. C. le Monnier, *Histoire céleste*; J. B. du Hamel, *Regiae scientiarum Academiae historia*; G. Bigourdan, *Histoire d'astronomie d'observations* (the foregoing eleven books presented by Captain R. S. Webb);

C. W. Allen, *Astrophysical quantities* (presented by the author); and

M. Waldmeier, *Ergebnisse und Probleme der Sonnenforschung* (presented by the author).

Professor H. G. Kienle, Professor O. Heckmann and Dr H. L. D'Azambuja were duly admitted Associates of the Society.

Papers were read by Dr A. H. Jarrett and Mr B. E. J. Pagel, and accounts were given, by Professor F. J. M. Stratton and Dr R. v. d. R. Woolley respectively, of papers by Dr M. W. Feast and Mr A. Przybylski.

Other meetings and activities in connection with the Society's visit to Glasgow included an Astronomical Colloquium on *The Formation of Stars*, held in the morning of September 7 and, on the evening of the same day, a Public Lecture entitled *The Evolution of Stars* by Professor Otto Struve of the University of California. Trips to places of interest included all-day excursions to the Trossachs and Loch Fyne; a visit, at the invitation of Mr Ian Garvie, to the optical works of Messrs Barr and Stroud; and inspections of the University Departments of Natural Philosophy and Zoology. An informal gathering was arranged for Fellows and guests by the Astronomical Society of Glasgow on September 7, and the proceedings ended on the evening of September 8 with a reception given by the University in the Bute Hall.

MEETING OF 1955 OCTOBER 14

Professor Sir Harold Jeffreys, President, in the Chair

The election by the Council of the following Fellows was duly confirmed:—

Kenneth Frank Bowden, The University, Liverpool 3 (proposed by J. Proudman);

Colin Stanley Gum, M.Sc., Department of Physics, The University, Adelaide, South Australia (proposed by R. v. d. R. Woolley); and

Jean-Claude Pecker, Laboratoire de Physique, 34 Avenue Carnot, Clermond-Ferrand, France (proposed by R. O. Redman).

Ninety-three presents were announced as having been received since the last meeting, including:—

J. B. Sidgwick, *Observational Astronomy* (presented by Messrs Faber and Faber, Ltd.); and

M. H. Jurdak, *Astronomical Dictionary—Arabic and English* (deposited by Professor H. Dingle, by permission of the General Secretary of the International Astronomical Union).

MEETING OF 1955 NOVEMBER 11

Professor Sir Harold Jeffreys, President, in the Chair

The President announced the death of Eugène Delporte, an Associate of the Society, and paid a brief tribute to his memory, the Fellows standing.

The election by the Council of the following Fellows was duly confirmed:—

Robert Hanbury Brown, Jodrell Bank Experimental Station, Lower Withington, Macclesfield, Cheshire (proposed by A. C. B. Lovell);

Douglas Richard Proctor Coats, 333 Kingston Crescent, St Vital, Winnipeg, Manitoba, Canada (proposed by R. Lockhart); and

Llewelyn Robert Owen Storey, 11 Manby Road, Great Malvern, Worcestershire (proposed by J. S. Hey).

The election by the Council of the following Junior Member was duly confirmed:—

Alex W. Rodgers, Commonwealth Observatory, Mount Stromlo, Canberra, Australia (proposed by A. R. Hogg).

Eighty-nine presents were announced as having been received since the last meeting, including:—

A. Unsöld, *Physik der Sternatmosphären* (presented by the author);

H. C. King, *The history of the telescope* (presented by Messrs Griffin & Co., Ltd.);

S. Tolansky, *An introduction to interferometry* (presented by Messrs Longmans, Green & Co., Ltd.);

Proceedings of the Astronomical photoelectric conference (presented by the Astronomy Department, Indiana University); and

Hamburger Sternwarte, *Katalog schwacher Sterne* (presented by the Hamburg Observatory).

MEETING OF 1955 DECEMBER 9

Professor Sir Harold Jeffreys, President, in the Chair

The election by the Council of the following Fellows was duly confirmed:—

John Olpherts Campbell Alleyne, South Lynch, Hursley, Hampshire (proposed by W. H. Steavenson);

Derrick Stanley Box, B.Sc., A.Inst.P., 63 Warwick Road, Thornton Heath, Surrey (proposed by A. D. Thackeray);

Donald Ross Brown, B.A., Iraq Petroleum Company, 214 Oxford Street, London, W.1 (proposed by G. Tait);

Robert Luke Bryant, 42 Centre Dandenong Road, Cheltenham, S.22, Victoria, Australia (proposed by W. G. L. Clapham);

Werner Buedeler, 101 Salisbury Square House, London, E.C.4 (proposed by P. Moore);

William Balson Caunter, Old Shortlands, Plaistow, Billingshurst, Sussex (proposed by D. W. Dewhirst);

Robert Clemitson, 1 Park Walk, Quarry Bank, Brierley Hill, Staffordshire (proposed by H. R. Didcock);

Raymond Davies, B.Sc.(Eng.), A.C.G.I., A.M.I.C.E., 10 Earlsway, Curzon Park, Chester (proposed by E. G. Williams);

William Harold Dickman, 411 George Street, Midland, Michigan, U.S.A. (proposed by D. H. Menzel);

Brian Emerson, B.Sc., 116 Middleton Boulevard, Wollaton Park, Nottingham (proposed by M. Johnson);

Yoshio Fujita, Department of Astronomy, University of Tokyo, 3-18 Azabu-Zigura, Minato-ku, Tokyo, Japan (proposed by W. Buscombe);

Heinz Gollnow, D.Phil., Commonwealth Observatory, Mount Stromlo, Canberra, Australia (proposed by A. R. Hogg);

Rev. Gregor Hagemann, D.Sc., Commonwealth Observatory, Mount Stromlo, Canberra, Australia (proposed by A. R. Hogg);

Ahmed Hammad, Professor of Mathematics, Faculty of Science, Cairo University, Giza, Egypt (proposed by S. Chapman);

Takeo Hatanaka, Professor of Astronomy, Tokyo Astronomical Observatory, University of Tokyo, Mitaka, Tokyo, Japan (proposed by D. H. Menzel);

Wulff Dieter Heintz, University Observatory, Munich, Germany (proposed by S. C. B. Gascoigne);

George Howard Herbig, Ph.D., Lick Observatory, Mount Hamilton, California, U.S.A. (proposed by R. H. Garstang);

Willem Daniel Kolkman, Pine Lodge, 19 Verulam Avenue, Purley, Surrey (proposed by C. R. Armstrong);

E. A. Kreiken, Professor of Astronomy and Director, Astronomical Institute of Ankara University, Fen Fakultesi, Ankara, Turkey (proposed by W. M. H. Greaves);

Maung Min Lwin, 58 35th Street, Rangoon, Burma (proposed by G. W. Wakeford);

David Jackson Mercer, Mimoza, The Crescent, Sassafras, Victoria, Australia (proposed by E. B. Walton);

Antoni Przybylski, D.Sc., (Tech.)Ph.D., Commonwealth Observatory, Mount Stromlo, Canberra, Australia (proposed by A. R. Hogg);

William Leslie Rae, 55 North View, Aspatria, Carlisle, Cumberland (proposed by J. L. MacDonald);

James Campbell Rome, Cdr., R.N., Naval Ordnance Inspection Department, Admiralty, Bath, Somerset (proposed by G. W. Ottaway);

Evry Léon Schatzman, Institut d'Astrophysique, 98 bis Boulevard Arago, Paris, France (proposed by M. J. Seaton);

Alexander Simpson, B.Sc., Marine and Technical College, South Shields, Co. Durham (proposed by F. J. Acfield);

E. Bennett Smith, 193 Allensbank Road, Heath, Cardiff (proposed by H. Bondi);

Francis Graham Smith, Cavendish Laboratory, Cambridge (proposed by G. Burbidge);

Stefan Szczyrbak, Kenwick House, Lords Lane, Heacham, King's Lynn, Norfolk (proposed by P. S. Laurie);

Arthur George Thomson, 148 Mill Lane, London, N.W.6 (proposed by E. Williamson);

Raimundo Oliveira Vicente, R Mestre Aviz 30, R/C, Algés, Portugal (proposed by H. Jeffreys);

John James Ward, A.M.I.Mech.E., 44 Blackacre Road, Theydon Bois, Epping, Essex (proposed by G. S. Brosan);

Joseph Witkowski, Ph.D., The Observatory, Poznan University, 36 Sloneczna, Poznan, Plodna (proposed by H. Spencer Jones);

Frank Bradshaw Wood, Cook and Flower Observatories, University of Pennsylvania, Philadelphia, Pennsylvania, U.S.A. (proposed by C. P. Olivier); and

Fritz Zwicky, Professor of Astrophysics, California Institute of Technology, 1201 East California Street, Pasadena, California, U.S.A. (proposed by R. v. d. R. Woolley).

Ninety presents were announced as having been received since the last meeting, including:—

H. P. Wilkins and P. A. Moore, *The Moon* (presented by the authors).

THE LUNAR SURFACE

T. Gold

(Communicated by the Astronomer Royal)

(Received 1956 January 13)*

Summary

Methods of classifying features of the lunar surface according to their relative ages are discussed. The origin of the lunar craters is thought to be meteoritic, and some new considerations are given concerning high velocity impacts. The investigation of relative ages suggests that a process of erosion has been operating, though of course much more slowly than on the Earth, and the hypothesis is then developed that eroded dust is subjected to an agitating action causing it to settle with the formation of flat, level surfaces. It is then unnecessary to introduce any hypothesis of lava flow for the generation of the flat surfaces, and a cold history for the lunar surface is then possible. The triaxial shape can then not be attributed to an equilibrium shape in a past epoch at another distance from the Earth, but merely as a shape that departs from the equilibrium by an amount compatible with the strength of the material.

---

1. *Introduction.*—Recent discussions of the lunar surface have mostly favoured the interpretation of meteoritic impacts for the generation of the craters, and of lava flows for the formation of the flat surfaces. There are two reasons for departing from the lava flow hypothesis: firstly, because there are some severe obstacles to that interpretation connected with the distribution of subsequent impacts on the plains and the absence of a satisfactory time sequence interpretation; and secondly, because it may be unnecessary to invoke both processes if the impacts alone seem sufficient together with other effects which are known to be present or which must be inferred in any case. The omission of the lava flow hypothesis is an attractive step also because of the absence or great rarity of lunar features that resemble the formations associated with terrestrial volcanic activity, although a far greater intensity of such activity has had to be assumed.

In the present discussion the arguments for the meteoritic hypothesis that have previously been advanced are considered so strong that it is not desired here to review that part of the evidence; but some additional considerations are given.

2. *The major surface features.*—The most obvious fact concerning the lunar surface is that there are two different types of ground. One type is rough, covered with multiple overlapping circular craters and comparatively light in colour. The other is darker, much smoother, and frequently so accurately level and flat as to resemble a liquid; and it possesses only a small fraction of the number-density of craters. The surface of the largest regions of flat ground —the maria—does not differ in appearance from that of smaller regions, and there is not much difficulty in classifying the surface everywhere according to these

*Received in original form 1955 August 19.

two types. We shall refer to them as "rock" and "filler". The "filler" is the material that has been regarded by many as solidified lava.

As has been pointed out by Baldwin and others, the craters are of shapes and configurations that would be adequately accounted for by intense explosions. Markings radiate out from some of them, the so-called rays, which can be reasonably interpreted as small debris flung out beyond the rim.

The craters, walled plains, and maria have usually been so classified according to the size of the formation. But, as is also pointed out by Baldwin, there is no real difference in type: it is a continuous distribution of objects without any clear breaks, although some features change gradually along the sequence. It is therefore undesirable to account for these objects by more than one basic process.

3. *The classification into a time sequence.*—It is useful first to investigate the methods whereby it is possible to classify features according to their relative ages. There is no process known which allows a classification according to an absolute time scale to be made; some dynamical considerations have been used, but they are indirect so far as detailed features are concerned and they have to rely upon estimates of the rate of transfer of angular momentum through tidal friction in the distant past, which is necessarily rather uncertain; and they have to rely upon the Moon having once possessed no stiffness and therefore having acquired the shape formerly appropriate to hydrostatic equilibrium, which the following considerations place in doubt.

There are various criteria for relative dating. The "criterion of overlap" is the most obvious and certain. If there are two craters that overlap each other then the one with the unbroken rim is the younger and the one whose rim was broken into the older. In the case of two craters that overlap and are of similar diameter the application of the criterion is entirely obvious. But it can also be applied in cases where one crater is very much smaller than the other, and is embedded entirely in the rim of the other. An explosion throwing up the large rim could not leave a small marking intact, and the small crater must hence have been formed after the large one. A large crater occurring after a small one in the same locality will lead to the total destruction of the small one, and such overlap cannot be seen except where the difference in size is not very great.

Overlaps of great multiplicity can be seen, and in some places it is possible to arrange five or six craters in a time sequence. (It is reassuring to note that while the criterion may be sometimes difficult to apply it never leads to logical conflict, such as that A is older than B, B older than C, and C older than A.)

Another relative age criterion is given by the rays. As these are features which must have originated at the same time as the parent crater and which overlie the surrounding surface it is possible to say that a crater with its rays is younger than any features over which the rays can be seen. Again this is a definite criterion that cannot reasonably be questioned in the framework of the meteoritic theory. But it is only of limited applicability owing to the small number of craters that possess visible rays.

The main importance of these two age criteria lies in the fact that they suggest and vindicate a third one: the sharpness of features. Crater rims and other detail vary greatly in the degree of sharpness, and it is found in all cases of overlapping groups of craters that when there is a clear difference between the members in this respect then it is the youngest one on the criterion of overlap which shows

the sharpest features. This rule can be checked with the aid of so many examples that it is not in doubt. The criterion of the rays implies in fact that the craters possessing rays are younger than any other features: the rays never appear to be in part overlapped by other features. It is therefore in agreement with the previous rule that the ray craters are also distinguished by the possession of the sharpest features.

The fact that the sharpness of features appears as an age criterion is not only of importance as a means of relative dating of formations that could not be classified by the other criteria; great interest attaches to it from the point of view of the physical processes that must for this reason be assumed to have occurred.

Instead of discussing the apparent sharpness and ruggedness of the rims of craters it is possible to include more measurable features without losing the significant correlation with the other age criteria. The height of the rim can be assumed to follow a certain law with crater diameter, as displayed by the youngest craters; the amount by which the rim of a crater is below the appropriate height can then also be taken as an indication of its age. Similarly, the loss of depth of the bowl is correlated with the other age criteria. But although the correlation of all these features with age cannot be doubted, it is clear that there is present a considerable spread for reasons other than age as well. The crater Copernicus, for example, which is 56 miles in diameter, has a rim only 3 300 feet high, whilst Tycho, with a diameter of 54 miles, has a rim 7 900 feet high; yet the possession of rays and the ruggedness of features and their sharpness in both cases excludes the interpretation of the difference as an age effect.

4. *The hypothesis of erosion.*—The age correlation of the sharpness of features of craters could be explained in two ways: either that they were of different appearance when formed, due to a progressive change in the process of formation or in the lunar surface material; or, and this is the explanation pursued here, that a process of erosion has taken place which has resulted in gradual changes. The interpretation using a process of erosion is made attractive because all changes that have to be accounted for are in the sense of diminishing the gravitational potential energy. The absence of any clear cut division in respect of the sharpness of features rules out any single sudden process of erosion, but implies that erosion must have been progressing through the major part of the interval of time during which the craters that are now visible were formed.

There is, however, no way of judging whether the rate of erosion has been constant or whether it has varied and perhaps diminished as the rate of bombardment diminished. Geological evidence and present-day terrestrial evidence places upper limits on the rates of bombardment sustained by the Earth and by implication also by the Moon. For an explanation of the lunar features it is undoubtedly necessary to suppose the rate of bombardment to have diminished very greatly since very early geological epochs. This would be in agreement with the point of view that the materials responsible for forming the bodies of the solar system have gradually been used up, leaving at the present time only a small fraction that has escaped capture. At the present time there may still be falling on to the Earth objects large enough to make a crater of the order of one mile in diameter at a mean rate of the order of one every ten thousand years. Such small craters do not stand up for long to terrestrial erosion and we can here only see the results of the last twenty or thirty thousand years; but even this small rate would not be insignificant on the Moon, where it would

imply that some tens of thousands ought to be visible unless they also suffered from erosion. It is unfortunate that there seems to be no consideration available at present that allows one to derive an absolute time-scale for the lunar events or to sort out to what extent the rate of erosion has been dependent on the varying rate of bombardment.

It should also be noted that there is no way of deciding how many craters may have been submerged by processes of erosion and bombardment; there is no hint that any surface is exhibited which antedates the meteoritic bombardment. Thus from an examination of the surface it is not possible to place a limit on the amount of bombardment that the Moon has received or on the depth of material which has been acquired by this process; and one cannot exclude the possibility that the entire Moon has been built up by the same processes that are now inferred from its present surface. If indeed the Moon had been put together in that way then it would not be unreasonable to think of smaller pieces having been put together earlier in the sequence of events; and therefore it would not be surprising to expect large impacts to have occurred in the process, so large that the present shape of the Moon has been dictated solely by the subsequent hydrostatic adjustments which the limit of strength of the material implies and not by any statistical symmetry which would have occurred had only finely divided matter been accreted.

If a similar history were envisaged for the Earth it would not follow that any of the major features here need still be attributed to the original bombardment. The Earth with its denser core and is lighter continents gives every evidence of differentiation of material having occurred and of internal changes having been a dominant factor in its evolution. The growth of continents and the folding of mountain chains many times over in the geological record are processes that have certainly been absent on the Moon; but on the Earth these processes have undoubtedly reshaped the entire surface.

The hypothesis of erosion has been accepted by many authors, in many cases without a discussion of the implications. To be satisfied with an explanation of the loss of height and sharpness of the crater features in terms of erosion it is, however, also necessary to find the present location of all the eroded material whose absence is postulated on some of the formations. We shall discuss later the possibility that the eroded material constitutes the "filler", and that this fills the floors of the maria as well as some smaller craters.

5. *The mechanism of explosions.*—In addition to the various arguments and considerations concerning meteoritic impact explosions and the craters so formed, which have been advanced by Baldwin, there are aspects of the physics of the problem that are worth noting.

Firstly, it is possible to draw a distinction between "explosions" and "bursts", according to the following definitions. An "explosion" is an event in which gas at a high pressure is suddenly made available in a locality, with the gas pressure far in excess of the maximum that could have been contained by the material there. A "burst" is an event in which there has been a source of gas under excess pressure, and where that excess pressure has risen to the value necessary to burst the containing material and break through to the surface. These processes are quite dissimilar, and so are their results. This is clear in the more familiar analogy of a boiler burst and a bomb explosion. The result of a steam boiler bursting due to the interior pressure rising above that which can be

contained, is always to produce a split, torn, and bent boiler. But if instead a powerful bomb were placed inside the boiler, whose peak pressure on exploding exceeded the maximum boiler pressure by a large factor, then the boiler would be torn into small fragments. The stream of gas would be too powerful to be channelled out through a few openings.

It is possible to judge the difference of appearance of a crater resulting from an explosion or a burst below the lunar surface. A burst would cause a steep walled, funnel shaped crater; and a more intense burst could only be caused by a source at a greater depth where a greater excess pressure could be contained, and from which would be formed a deeper funnel-shaped crater. But a very powerful burst could not have its origin close to the surface, and the mechanism could therefore not be used to explain the generation of a crater shaped like a wide and very shallow bowl.

It is quite different with an explosion, for there an arbitrarily strong explosion can be considered at any depth including at the surface itself, where it will generate a shallow depression. The great majority of the very large number of ordinary lunar craters ($\sim 10^5$ or 10^6) require an explanation in terms of powerful surface explosions and not of bursts.

It is worthwhile to consider qualitatively, but in more detail, the physical mechanism of a meteoritic impact and the type of explosion resulting. The process differs from the one of the impact of a rifle bullet in the essential respect that it is highly supersonic: the velocity is far in excess of the speed of sound in the colliding materials. A meteorite at 50 km/sec is thus at a collision Mach number of about 10, while rifle bullets or shells are usually at collision Mach numbers of the order of $\frac{1}{2}$ or $\frac{1}{4}$. For the supersonic collisions, a new type of process must set in which is absent for the more familiar infrasonic ones: a shock wave must be generated in both the colliding solids, whose wave front is sufficiently compressed for the resulting rise in temperature to increase the speed of sound, initially up to the impact speed. Before this has happened there is no method for the materials to get out of each other's way, and the pressures must hence rise to reach that condition.

For large values of the Mach number it is legitimate to ignore the solid state binding energies; apart from small factors the thermal particle speed has to be raised to the impact velocity, and that has to occur in a mass of the same order as the mass of the missile before the speed is reduced so as to halve the kinetic energy. This implies that the bulk of the kinetic energy is converted into heat in a layer only about twice as thick as the missile (assuming a similar composition of missile and ground). The order of magnitude of the temperature at which materials such as may compose the lunar surface possess a speed of sound of 50k/msec is one million degrees, and such temperatures will therefore be reached in the initial stages of impacts occurring with these velocities.

A fraction of the missile is likely to be still solid when decelerated to the region of Mach number $M = 1$, and this will have a different behaviour. It will dig itself in, being able to dissipate the remainder of its energy only comparatively slowly along its path by the processes of fracturing the solid, by the soundwaves radiated from the path, and by the friction (perhaps the main effect) between the sliding solids. The very much more rapid dissipation of compressional heating is confined to the high Mach number part of the impact trajectory. The depth of penetration before $M = 1$ is reached will vary approximately as $L \log M_0$ where

L is the thickness of the missile and M_0 the initial Mach number (assuming similar materials throughout), as it would if the loss of momentum to the shock was the only relevant effect.

It is a consequence of these considerations that an object would penetrate into a solid body less far for having a higher initial speed, once the speed is above a certain value. The contribution to the penetration provided by the supersonic part of the trajectory cannot be made more than a small number of diameters of the missile; but the mass, strength and shape of the object setting out on the subsonic penetration will matter greatly to the final penetration depth. The destruction of the missile in the supersonic region implies that a diminished mass will commence the subsonic travel, and for that reason penetrate less far.

In a meteoritic impact (at $M \gg 1$) the bulk of the energy will be available in the form of an extremely hot gas (a million degrees or so) within a distance of the original surface of about two diameters of the meteorite. This means in effect that a surface explosion takes place, for the energies are high and the extremely hot mass of gas will expand rapidly and affect severely regions that are very large compared with the size of the meteorite . This also indicates that the main effects of the explosion will be similar to those from a point source, and that therefore the initial direction of the meteorite will not have much influence on the symmetry of the larger markings left by the explosion.

The shallow wide circular bowl shapes of the lunar craters are thus in agreement with an explanation based on the considerations given, and so are the shapes of the well preserved meteoritic craters of Arizona and Northern Quebec (Chubb Crater). Also it is to be expected that the remains of the meteorite should be well below the bottom of the explosion crater, having penetrated subsonically into the ground, and that their location should be given by the initial direction of motion, while the pit blown out by the hot gas is nearly symmetrical around the impact point. This agrees with the findings at both these terrestrial craters.

It is desirable to contrast this picture with the one that would result from any form of internally produced "bursts". No reason has been found there why an increase in the scale of the phenomenon should lead to a change in the geometry, and the shallow, wide bowls which are characteristic of the lunar crater shapes would remain unexplained. Theophilus, for example, is a representative crater in which there is no evidence of any change after its formation; the diameter of the rim is about 100 km and the total depth of the bowl is only $4\frac{1}{2}$ km. The release of any excess pressure that could have been contained at such a depth is clearly quite inadequate as an explanation of the main crater. Suggestions of any intrusive or volcanic action as the origin of the majority of lunar craters are also inadequate to explain why the formation of craters high up in the mountains thrown up by previous events should be as common as elsewhere.

It is now possible to discuss in more detail the shapes of craters to be expected from various impacts. For a given initial velocity of a meteorite of linear dimensions L the effective depth of the explosion caused varies as L, and the energy released as L^3. The diameter of the crater caused may be expected to correspond to the diameter at which the explosion wave spreading out over the surface has diminished its energy density to a certain value, and it would thus vary as $L^{3/2}$ The diameter of the bowl thus goes up faster than the depth of the explosion with increasing size of meteorite, and the whole process does not scale up with geometrical similarity but becomes comparatively more a surface phenomenon.

We may suppose that the final shape of the crater bowl resulting from an explosion is determined mostly by the actual surface transport of material and not by any permanent effects of the brief pressure wave that swept over it; (for owing to this brevity the attenuation with depth was very great and only a thin layer was subjected to extraordinary pressures). An explosion cannot remove any material from directly below its origin, except in the immediate neighbourhood where the wavefront is still complicated and irregular owing to the finite size and irregularity of the source. There must exist in any one case a certain "angle of attack", defined as the greatest angle to the upward vertical of the direction in which material is set in motion by the explosion and eventually expelled (Fig. 1). This angle of attack will depend upon the magnitude of the

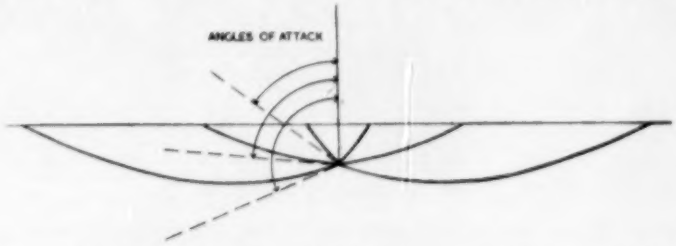


FIG. 1.—*Angles of attack for a submerged explosive centre. The lines represent the limit down to which material is scooped out by the explosion. Greater explosive power is expected to result in larger angles of attack.*

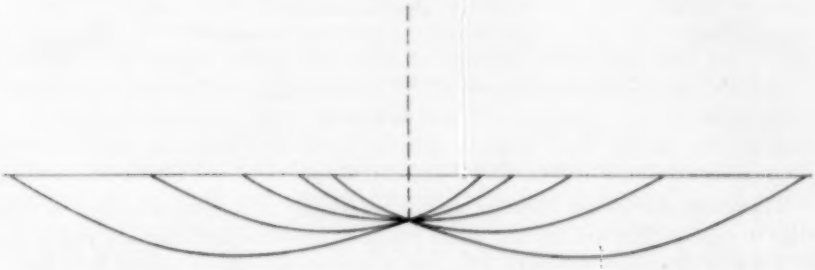


FIG. 2.—*Illustration of the variation of crater shape that is expected with a variation of the angle of attack. Schematic, vertical scale is exaggerated.*

vertical and horizontal components of the force applied by the gas to the material, and on the properties of the material. The process may be susceptible to theoretical treatment, but not in any simple manner; experimentally these effects are known for the case of bombs exploded at a shallow depth in terrestrial ground, but it is doubtful whether any artificial explosions have been powerful enough to exhibit a large angle of attack, of the order of 90 degrees or more (leaving aside the immediate neighbourhood where the finite size of the source matters). The possible patterns of streamlines of the flow of the solid suggest that for angles of attack of more than 90 degrees a very great increase of the force is required for a small further increase of the angle, for the amount of material scooped out and the diameter of the crater formed must increase very rapidly with this angle; yet these quantities must maintain a reasonable relationship with the primary energy available (Fig. 2).

The steep funnel-shaped holes of weak explosions originating some way below the surface has already been referred to in the extreme case of "bursts". There the angle of attack is small. As the strength of the explosion is increased (assumed at the same depth) the angle of attack increases causing wider bowl-shaped craters; and when the angle of attack has risen to more than 90 degrees the resulting craters will no longer be deepest in the middle but at some distance out from the middle. A qualitative consideration of the possible family of shapes makes it clear that a further increase in the strength of the explosion must increase both the maximum depth and the distance at which it occurs.

The stream of the scooped-out material and gas must gain upward momentum so long as it is exerting a downward force on the ground in excess of its gravitation. The initial downward motion must therefore eventually give way to an upward one. When the stream then rises above the level of the original surface the condition for the support of the scooped-up solid is changed, for it can no longer derive gravitational support from the ground below at the same time as moving with the gas. It may then settle out of the stream of gas, falling below the lowest streamline and thus into a region which, in the circumstances, must be entirely sheltered from the gas. Dust particles and small pieces that could acquire a high speed from the gas will be projected furthest (Fig. 3).

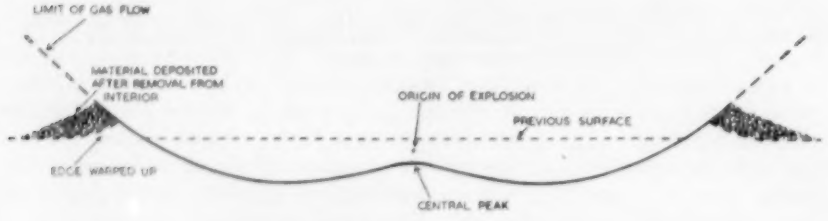


FIG. 3.—*Interpretation of the features of a crater.* (*Vertical scale exaggerated.*)

These considerations allow of an interpretation of the main shapes of the range of craters that are not too much spoilt by erosion (and that are young on the basis of the other criteria). Beyond a certain size these craters frequently possess a central peak, suggesting that for that strength of explosion the angle of attack is in excess of 90 degrees. But even for the biggest explosions whose markings can still be clearly seen the angle of attack is only about 15 degrees greater than a right angle, so that 15 degrees is then the steepest slope of the central peak, occurring very near the summit. It is necessary, on this explanation, that the central peak should never be higher than the terrain surrounding the entire crater, but always somewhat lower. This is indeed so. In the case of unspoilt craters in flat terrain the depth of the central peak below the plain may be taken as an indication of the size of the responsible meteorite; with a high speed meteorite, and an allowance for some vertical digging due to non-central forces, one may estimate that this depth corresponds to between four and six meteorite diameters. For very large craters all these considerations may be systematically in error due to the isostatic adjustment that must be expected immediately following the explosion; and for small craters the chance effects of lack of symmetry of the explosion must proportionately play a larger part.

The shape of the inside of a crater bowl is determined by the limiting streamline of the flow. But the outer side of the crater wall is shaped by the fall-out of the solid pieces and therefore by the velocity distribution of those pieces during their motion (and to a smaller extent by the upward warping of the edge). (Fig. 3.) Neither surface will therefore be expected to develop a steepness equal to the angle of repose of the broken-up material. The absence of such gradients has sometimes been remarked upon in discussions of theories of origin of these formations.

Most craters on the Moon are remarkably close to circular, and this could not be expected from explosions whose origins extend over a large part of the crater. A very high intrinsic explosive power is indicated in order to assure that the blast is effective at a range great compared with the size of the explosive object; this will then act as a point source whatever its shape, provided it is reasonably compact. High velocity impacts causing a conversion of a mass of the order of that of the missile into a gas at a million degrees or so (compared with merely some thousands of degrees for a chemical bomb) will provide such a high intrinsic power, and therefore blast a pit whose shape is independent of the shape of the meteorite or the geometry of the impact (except in extreme cases).

If any infrasonic impacts have occurred on the Moon they must have left entirely different markings from the usual shallow circular craters. The missile would possess little explosive power, and could in any case not generate gas at a temperature of more than a few thousand degrees. The missile would dig itself in and leave a steep-sided hole whose shape and direction depended upon the shape of the missile and the geometry of the impact.

Even without a knowledge of the detailed mechanism of high velocity impacts the great change in the type of marking produced in changing from an infrasonic to a supersonic impact can therefore be recognized. The great majority of the lunar craters demand an explanation in terms of supersonic impacts.

Even more detailed features can be interpreted now. It has often been remarked that the central peaks frequently show a small pit at their centre. Baldwin dismisses this as being no more than the expected density of small craters on any ground: but one can criticize the statistical argument used, as no allowance is made for the fact that the peaks showing the phenomenon mostly belong to craters that appear to be very much younger than the majority, and many younger than any except ray craters (possibly because this detail is soon lost by erosion). If the probability of subsequent crater formation only is used, then the central pits appear much more significant. The absence of pits of similar size in the remainder of the bowl and rim of these craters is significant, for this can be simply regarded as a representative area registering only later impacts. (Plate 9 shows a group of craters with central peaks and pits.)

This detail is fully expected in the case of supersonic impacts. The source of the explosion is distributed along the path of the missile but tapers off steeply in intensity, until after three or four diameters of the missile from the contact point the remaining piece of missile ceases effectively to produce hot gas, but continues to force its way through the material at subsonic speeds. A steepening funnel must therefore be produced, as the explosive power diminishes and the penetration continues (Fig. 4). This feature would be steeper than any of the other parts of craters, and might therefore be expected to erode or fall in more quickly. In the case of the smaller lunar craters this feature would be too small

to be seen. But in the case of the Arizona crater it is known that a remainder is buried deeply and to one side of the centre. This, by lunar standards, is a very small crater and therefore does not possess a central peak; but the central pit must have been there, at the beginning of the tunnel leading to the buried mass. It will not have been quite in the centre owing to the skewness of the impact, an effect of greater importance for smaller craters. By now all this is silted over as is the entire bottom; but there appears to be little silt in the case of the bigger Chubb Crater, and it may yet be possible to find the central pit there under the water that has filled the lower part of the crater.

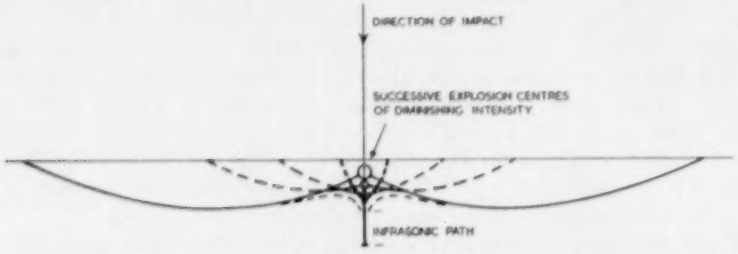


FIG. 4.—*The formation of a central peak and a central pit. Successive centres of explosion are shown, each with the crater that would have resulted. The actual crater resulting from an explosive line source whose power diminishes rapidly with depth may be expected to resemble that obtained by superposition of many point sources.*

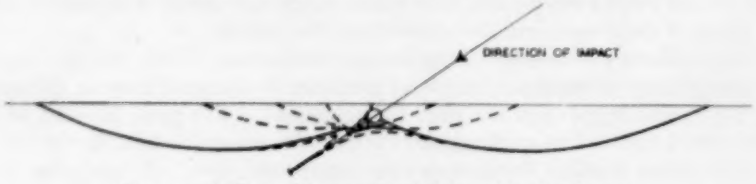


FIG. 5.—*As Fig. 4, but central peak deformed by shew impact.*

The entire complex shape of a crater with a rim, a central peak, and a pit within that is thus attributed to the single effect of a diminishing angle of attack with diminishing explosive force. But it will be seen that the effect of the central peak and its pit is very sensitive to departures from symmetry of the explosion. Particularly non-vertical incidence will have an important effect on these structures, while the large bowl and rim of the crater caused by the main force of the explosion, derived in any case from the close vicinity of the impact point, are not significantly affected. In the case of a very inclined track (Fig. 5), successive centres of the explosion are laterally displaced and the central peak that would be left by the initial centre of the explosion is vulnerable to the weaker but displaced later stages. Baldwin remarks that the central peak is a feature that appears to lack much correlation with the other features of craters; and this would be accounted for by the wide distribution of the angles of incidence of meteorites.

6. *The flat plains.*—On the material referred to as "filler" the number-density of craters is very much smaller than on the "rock". With a meteoritic origin of craters it is necessary to account for this distribution by some process that occurred after the formation of most of the craters. The "filler" material must have been distributed over the low ground in the characteristically flat

Region around Copernicus, showing craters possessing a central peak and a pit within that. 100-inch Mount Wilson.

Full Moon showing ray structure of the prominent ray craters Copernicus and Tycho. Chappell, Lick 36-inch refractor.

and level manner, drowning or obliterating all existing craters in these regions and registering subsequent impacts only. Lava has usually been considered in this connection.

There are several objections to the lava flow hypothesis, which probably owed its origin to the connection with volcanism when a volcanic origin of the craters was considered probable. The most important objection is connected with the impossibility of finding a plausible time sequence for the events.

The filler fills the largest craters, the maria, as well as many others of much smaller size distributed without obvious pattern over the rest of the hemisphere. The density of craters made in the filler material is low in all areas of filler. On the basis of the lava flow hypothesis it must therefore be supposed that lava flows occurred at about the same epoch and that this was at the stage when only a few per cent of the visible bombardment was still to come. The rims of the maria and of the craters filled with filler show, however, the highest density of small craters that occurs anywhere. It could therefore not be supposed that the process of formation of these large craters was itself responsible for letting in the lava by establishing a hole in a supposedly thin crust. The lava would have had to find its way into all those craters separately much later than they were formed. This process would have had to occur for many comparatively small craters on high ground as well as the large ones and the connected maria. Yet no independent signs of break-through can be seen on all the large regions of connected high ground, most of whose craters antedate the supposed floods. Although supposedly frequently punctured in the bottoms of craters, the ground never cracked to produce intrusions and out-pourings of lava from fissures. Any substantial distortions of the ground could easily be recognized from the shapes of the old craters, and such distortions are conspicuously absent. There is no buckling, no overthrusting, no folding, yet old craters are frequently filled up to some level, often quite a different level in nearby but unconnected craters.

The requirement would therefore be for a very large amount of melting to have occurred on the Moon in a late phase after the bulk of the bombardment was over. Melting would have to occur at such shallow depths that even comparatively small craters that had long existed on high ground would let the lava in. The lava must therefore have been within 3000 feet or less of the surface in very many places. A large amount of hydrostatic unbalance, both due to the great height of existing crater rims and also implied by the different levels of filling, would, however, have to avoid causing any distortion of the large regions of the order of a thousand miles of connected highlands. This is not plausible. Alone the changes of mean density associated with melting or compacting would have produced bigger effects; and the out-pourings on the enormous scale and the high fluidity for the lava to run flat over very large distances cannot have left a thin skin of broken-up material undisturbed, except for many local intrusions into the bottoms of shallow depressions.

Though the connected maria have often been assumed to be the result of a single out-pouring there are many craters in the outer regions of these maria that have unbroken rims standing high above the level surfaces, and that are nevertheless filled with filler. Each one of those must have been punctured from below separately. This hypothesis is clearly unsatisfactory.

The great regions of quite undisturbed highland, the absence of large volcanic structures, of foldings and large rifts or thrusts over periods that must be

comparable with or longer than the entire interval of terrestrial geology, suggests not only the absence of volcanic events huge by terrestrial standards; but it suggests that the Moon has been mechanically very much more stable than the Earth and that the stability must be sought in lesser effects of heating and the attendant changes of arrangement and of density of materials. These considerations favour the view that the Moon has been substantially cold throughout the period registered by any of its present surface features.

It has been suggested (Urey, 1952) that the lava could have been produced by the heat supplied in the major impacts. This suggestion also runs into the difficulty concerning the much later date of formation of the flat surfaces than of the rims surrounding them. Also it is considered very unlikely that large quantities of liquid can result from impacts. Liquefaction would have to be due to the transport into the material of energy made available at the impact. The transport by conduction of heat cannot account for more than a very thin layer of liquefaction. However much heat is supplied externally to a piece of solid material, its surface cannot be maintained at a temperature higher than the one at which it vaporizes. (If heat is supplied at a higher temperature, then the rate of vaporization may be high and the vapour may heat up to the high temperature.) The depth below the vaporization surface to which enough heat for melting can be conducted in the brief interval of the order of seconds during which heat is supplied in the case of an impact explosion, is given by the material properties and does not exceed some centimetres for any material. If the amount of heat supplied is much in excess of that necessary to keep the surface at the vaporization temperature, then the amount of melting will be diminished, for the vaporization front will penetrate into the material, chasing the liquefaction front.

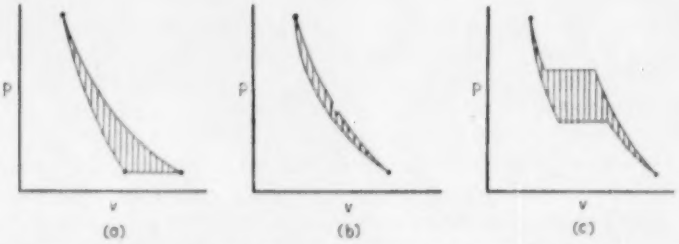


FIG. 6.—*Pressure–volume diagrams for heating by a single pressure cycle.*
(a) When the final density is different from the initial.
(b) With hysteresis, but final density same as initial.
(c) When there is a pressure phase change whose critical pressure depends upon the sense of the pressure change.

The other possible mechanism of distribution of heat is by the shock wave progressing through the material. A single pressure impulse of adequate intensity can cause melting, though special material properties are required if vaporization at the maximum pressure is to be avoided. The most favourable material for this will be one whose final density at the end of the pressure cycle is much greater than before the application of the pulse, and where consequently a large change in the internal energy is implied (a large area in the pressure-volume diagram, Fig. 6). Sand would be a material of this type. Whether the heating resulting from a single pressure pulse can ever be large in a material whose final density is similar to the initial one is uncertain; the energy available there can

only correspond to the area between the two curves of the pressure-volume diagram, whose end points are common. The effect depends upon the material possessing very much hysteresis in the pressure-density relation. It is possible, of course, that pressure-induced phase changes may show substantial hysteresis of this sort.

In the case of an explosion the sound wave radiating into the material suffers geometrical attenuation so that the energy density diminishes as the inverse square of the distance from the focus, augmented by the high physical attenuation implied by the intended melting. The energy density must therefore diminish rapidly from the focus and the heating would be expected to diminish even more rapidly. This will restrict the volume in which the heating is adequate to melt and inadequate to vaporize. It is to be expected that in practice most of the energy will go into making a very hot vapour, and most of the remainder into fracturing and pulverizing the solid, but very little into melting, a process that requires too high a degree of uniformity in the conditions and too small a range of temperatures at which heat has to be supplied.

In connection with the lava flow hypothesis the view has been advanced that the darker colour of the maria compared with the high ground is simply the consequence of the lava being a darker material than the other type of rock. This view leads to insuperable difficulties. If erosion is assumed to be the cause of the rounding of the features of old craters, as discussed earlier, then it is necessary to find the eroded material. This cannot reasonably be expected to have all left the Moon entirely. If an estimate is made of the amounts missing from the rims of all the many overlapping old craters, then this cannot come to less than the equivalent of a 300 foot depth if it were distributed over the entire surface, and probably a great deal more. Though on the lava flow hypothesis it could be supposed that the eroded material was so distributed that it was all drowned, the question would remain as to the whereabouts of the material eroded after the supposed lava flooding. This is still a substantial depth, at any rate deeper by a large factor than is necessary to make it optically opaque. Even if the erosion process were not to take the erosion product any further than is strictly required by the change of shapes, there would still have to be large areas of the flat plains covered over by that material. This is clearest in the cases of flat-bottomed craters whose erosion on the inner crater walls must be assumed to have covered over at least part of the flat bottom, yet the colour there even close to the sharp edge is the same as that of the flat surfaces elsewhere. A more diffuse distribution of the eroded material would only strengthen this argument, so it must be concluded that the darker shade is that of eroded material even if it came from lighter rock.

The disappearance of rays demands a similar explanation. If older craters also produced rays as the most recent ones have done, and as must indeed be expected, then the combined ray material would entirely cover all the surface with much more than optical thickness. On the high ground the occurrence of erosion will itself account for the removal of the ray material, but what has happened to all the rays that ought to be on the flat ground due to craters subsequent to the supposed lava flows? There are quite enough of those to have covered over the majority of the areas of the maria. All this material must either have changed its colour and become darker or been covered over with eroded material from elsewhere which must in any case be presumed to be dark. In no case could the original lava surface be the one that is visible at present.

It may also be presumed that in the so called post-lava period when still a few per cent of the bombardment came this would have been associated with enough dust to cover over the entire surface opaquely. Indeed, the estimates of dust at present still being acquired by the Earth (and Moon) would assure such a cover in a comparatively short time. Petterson and Rotschi (1952) estimate from the nickel content of deep ocean deposits that the quantity of material currently deposited on the Earth is of the order of one million tons per year. This estimate would imply that the Moon is acquiring a layer one centimetre in thickness every 10^7 years, at the present time. As this contribution is in the form of a very fine powder the individual impacts cannot cause any disturbance at a depth greater than about one millimetre, and except for areas of larger impacts there can be no significant mixing of materials with those at greater depths. Therefore an opaque blanket would be spread over the entire surface, and renewed once every million years. (This consideration would be invalidated if in each little impact an amount of vapour was freed and allowed to escape from the Moon entirely, of a mass of the same order as or exceeding that of the micro-meteorite; or if a fraction escaped the period would be lengthened. The material making up the rays of craters would have to be at least some centimetres thick if it is not to be covered over in too short a time to account for the absence of similar craters on the Earth. Until the rays are submerged they would have to be able to shed dust just like the highlands.) A process of migration of dust from high ground to the lowlands can again be invoked to account for denudation of the high ground and for that reason for a differentiation in colour. Again the actual surface in the flat regions would have to be dust, and nothing is gained by supposing lava to have been darker than rock.

If erosion has been continuous then it is necessary to suppose the present ray craters to be very young compared with the age of the great majority of the features. This will no doubt mean that they occurred in the time covered by geology on the Earth, while most of the other features can be considered older than the available geological record. A few similar stragglers hitting the Earth within the last hundred million years may well be expected to have their markings obscured by the vigorous processes on the Earth's surface, which must of course also be expected to have entirely obliterated a multitude of lunar type craters in perhaps three thousand million years.

The ray craters need not be so young if the processes of erosion and changes of colour of the surface were dependent upon the infalling material. In that case the great diminution of the bombardment which must in any case be inferred will have implied a similar diminution in the rate of erosion, and the last craters to have been formed must be uneroded.

The amounts of eroded material missing on the high ground are large, and the absence of any signs of this material on the filler, especially around the sharp edges of maria, or the sharp junction between crater walls and flat surfaces, requires in any case a special hypothesis whatever origin of the flat surfaces is assumed. It is necessary to suppose that the eroded material distributes itself over the flat surfaces. This supposition disposes at once of the difficulty concerning the disappearance of rays on the flat ground. These very thin markings would merely be covered over. If, then, it is necessary to suppose that a mechanism exists that allows the eroded material, presumably very fine dust, to "flow" like a fluid, then it is no longer necessary to invoke any other fluid ever to have

been present to cause the flat regions. We shall therefore pursue the consequences and requirements implied by the hypothesis that all the flat regions are beds of dust whose origin lies in the erosion of the high ground as well as perhaps in newly acquired meteoric material.

7. *The flow of dust.*—The smallness of the gradients in most of the sediment and the sharp distinction between the eroding rock and the sediment suggests that the flat distribution must not be the process that limits the speed of the erosion. The requirement is for the transport of dust particles to the minimum height that they can reach to occur at a rapid rate compared with the other processes of the formation or freeing of dust at the surface or its acquisition from outside. Compared with terrestrial standards the transport requirements are, of course, extremely low: fine dust particles on the surface must be moved only at such a speed that the maria can be filled to an average depth of perhaps a thousand feet in a period that may be three thousand million years; one millimetre in ten thousand years.

The requirement is for the average speed of flow not to diminish appreciably with the angle of the slope until a very small angle is reached. It should not be a flow restricted by the equivalent of viscosity in a liquid, for this would leave slight gradients persisting for long and remove steep gradients very quickly. The type of flow required would be one where at any instant a thin layer on the surface behaves like a non-viscous liquid whilst the remainder underneath remains stiff. This would then allow steep slopes to persist but would assure that the deposits of dust possess a flat surface. Such a "fluidization" of a surface layer would have to be a process resulting from an external energy source, providing an agitation for the dust particles on or near the surface.

There appear to be some very steep slopes in the filler material. In particular there are the small rills which possess very steep sides. This is not contrary to the known behaviour of dust. At a depth of more than a few metres the compaction under the overlying weight would suffice to convert the material effectively into a solid; in the absence of an atmosphere grains stick together with intermolecular forces with even less compaction than is necessary when atmospheric molecules intervene. If the material can be regarded as resistant to internal flow but possessing a thin fluidized surface layer, then the filling up of steep holes or gaps would proceed at a speed limited by the fluidization process, but the steep sides would persist until the filling up is complete. It is not a requirement that such rills should survive for a great length of time; Baldwin speaks of them as the youngest features on the Moon, yet has to add that some appear partly "filled with lava". These rills are to be thought of within the framework of the present interpretations as the small signs of the small isostatic adjustments that keep occurring as the maria fill up with dust and the highlands are denuded; and these rills are in turn comparatively quickly filled in. Their distribution mainly around the edges of the great maria is entirely in accord with this interpretation.

The "fluidization" process remains to be discussed. It would, of course, require to be discussed in any other interpretation of the lunar surface, provided that the worn appearance of the old craters is attributed to erosion; the only difference here is that the process needs to be effective enough not only to transport the minimum amount of material that is certainly known to have been *removed*, but the somewhat larger amounts that may have been removed or acquired from

outside and which are certainly known to have *accumulated*. But in no other way is the requirement changed, not even in the degree of flatness of the end result of the process.

The correct choice of the physical process responsible for the fluidization is a matter of the greatest difficulty; but it is not difficult to suggest processes that may be responsible and for which it cannot be said that they must fall short of the requirements. Perhaps the most serious consideration should be given to the effect of an evaporation and condensation cycle of a suitable vapour, to the effects of micrometeorites, and to the effects of electrostatic forces. There may be several other possible processes. Much dust may have been moved by the gas produced in the largest impacts; a flat deposition is, however, not to be expected from a few violent jerks, but only from statistical effects of a process in which each particle was lifted up many times.

An evaporation reaction is exerted on a solid particle when a volatile material is vaporized from its surface. This force arises if vapour has condensed on to the particle and then is removed again, such as by the action of sunlight heating the particle. The order of magnitude of the effect is such that the condensed material has to be about one thousandth part of the mass of a dust grain in order that the grain may be lifted by one centimetre in fortunate circumstances. But a much more effective process may be occurring. A substance of such a molecular weight as to remain bound to the Moon in the vapour phase (molecular weight more than about 50), may condense in the shade and evaporate in the sunlight. Assuming a very small amount of atmosphere on the Moon only, the substance would remain concentrated near the morning edge; it would evaporate from the material moved into the sunlight and much of it would recondense in the nearby dark regions. The remainder would distribute itself over the sunlit hemisphere but would again be caught as soon as it moved into the dark hemisphere. A given quantity of vapour would be effective many times in one lunar day if the average distance of the condensation from the region of dawn is small. (The details of the process would depend very much upon the amounts of other gases present, but for the present purposes only in such a way as to affect the quantities required for the dust transport mechanism.) Before condensing, the vapour may be expected to sink into the interstices between the dust particles and when the region becomes heated the substance will be vaporized and driven out. This may be quite enough to agitate the layer into which diffusion had occurred, a layer of perhaps a few millimetres or a centimetre in thickness.

Another source of agitation of the surface is due to all the small meteoric dust particles that must be expected to strike. This mechanism would be inadequate at the present rate of bombardment, but the entire meteoritic theory would also be inadequate at that rate. A very much higher rate has to be supposed for the past and it is then not certain that all the small explosions due to particles in the range of 10^{-2} cm to 10^{-4} cm, with the puff of very hot gas generated in each one, cannot together have moved the surface dust. If the quantity of meteoric dust accumulated is thought of as being of the same order as the eroded dust, then such a method of transport is very likely. The process would satisfy the requirement of effectively fluidizing a surface layer only.

Electrical forces are also a possibility. These can arise either as a consequence of photoemission of electrons from the surface due to the ultraviolet light of the Sun, or they could arise in some larger electrical process connected with solar

events of the type that cause the aurorae and magnetic storms on the Earth. The process of photoemission in the absence of an atmosphere, and especially of moisture, will result in an erratic distribution of charge on the irradiated surface which may be a very good insulator. While the average positive charge will be inadequate to lift particles off, it is not clear that the chance distributions in small localities could not do it. (This type of mechanism could best be investigated experimentally.) If particles were frequently dislodged by such electrostatic forces, then again a net flow would result on the surface in accordance with the requirement.

The patterns of current induced in the Moon by the passage of a solar corpuscular stream have not been investigated. In the shadow zone the conductivity of space would be very low and electrostatic forces may therefore be of importance. Effects of this type may do more than merely fluidize the surface; they may actually have the tendency to lift up and scatter any dust above a level surface, thereby speeding the establishment of level surfaces.

If very small dust particles are the rule, then their thermal agitation may be adequate to cause transportation. Professor P. Morrison has suggested (privately) that such thermal agitation can be considered as resulting in an "atmosphere" whose molecules are the dust particles. On account of the large "molecular weight" the scale height would be very small. Such an atmosphere would have just the required properties; it would flow down even very slight gradients and would settle material in flat level surfaces if more material was made available from higher ground.

If the Moon ever possessed an atmosphere, then the transport of dust could of course just be ascribed to wind erosion. On the whole the appearance does not favour that interpretation, for there are no large-scale patterns and preferential directions of drifts to be discerned, while those would be expected from an atmosphere that derived its convective motion permanently from the temperature difference between the light and dark hemispheres.

8. *The generation of dust.*—A process of erosion generally consists of the two stages of the break-up of the material and of its transportation. We have so far considered only the one of transportation, and this may indeed be all that is necessary. The material of the lunar surface may be composed of loose aggregate with a large proportion of dust, especially if it has always been cold and if no substantial processes have been present to deliver compacted or fused material to the surface. All the impacts that are indicated would have generated much fine powder, and even if more substantial pieces are present in the crater walls they may be so embedded in dust that the removal of the dust would make them tumble.

It is, however, possible to suggest processes that would cause more dust to be formed from solid rock provided it is quickly removed and does not protect the surface. The solar radiation in the region of ultra-violet and soft X-rays will be absorbed on the lunar surface and may cause there a destruction of a thin surface layer. It must be remembered that all the radiation which is responsible on the Earth for the heating and ionization of the outer atmosphere will there be absorbed in a layer probably less than one millimetre in thickness. Equilibrium between the resulting crystallographic destruction and the spontaneous recrystallization of the material may occur only when a large proportion of fine powder has been formed. In addition, there is the possibility of escape from the

surface of certain constituents of the crystals. Oxygen, for example, could be released through the irradiation and the remaining material may not re-crystallize. Even if the same material had formed into crystalline particles at a similar distance from the Sun it is not necessary to suppose that those would be stable when later exposed on the lunar surface where the temperature reaches much higher values. The rates of such surface destruction required on the Moon to be of significance are only of the order of one kilometre in 10^9 years or 10^{-4} centimetres per year. There may also be a comparable destruction of the surface by fast impinging gas molecules in a manner very similar to the destruction of cathodes of vacuum tubes by positive ion bombardment. Lastly, micro-meteorites could be held responsible not only for the supply of new dust but also for the generation of dust from existing solids. Thus there is no difficulty in supposing an adequate supply of dust to be present. Possibly the initial freeing of dust particles may be a restricting factor in the erosion process if the particles are at first more tightly packed than they are once they have been moved.

The darker colour of the beds of sediment has to be explained. As mentioned earlier, the explanation has to be found in any case, whatever theory of the surface features was adopted, so long as erosion played a part. It is necessary to suppose that dust which has been on the surface for a certain length of time becomes dark; this might be merely due to its mixing with fresh meteoric material. But a more complicated behaviour is required. Even small craters in the maria show a lighter colour and it must therefore in any case be supposed that the dark colour is confined to a thin surface layer. Unless the colour of recent meteoric dust is different from the older variety, one has to suppose that some chemical action takes place which alters the colour of dust under the surface back to the lighter shade. Again the required rates are very low.

Another possibility would be that the darker colour is produced by the action of the solar X-rays on material that has been on the surface for a comparatively long time, whilst the denuded highlands keep presenting material that has until recently been protected. Such radiation-induced coloration is generally lost by heating, and this would explain why the fine powder that has been involved in an explosion is again turned to the lighter shade.

9. *Comparison of highland craters with those placed in maria.*—The comparison of the heights of the two "recent" craters, Copernicus and Tycho, has been mentioned. The much smaller height of the rim of Copernicus may be attributed to the fact that this crater was formed entirely in dust, and that dust may behave very differently from more compacted or rocky material in the explosion. There is also a very striking difference in the appearance of the rays of those two craters. Tycho's rays are very long and straight, whilst those of Copernicus are much shorter and fuzzy, and a much larger area near the crater has been covered with ray material. Again this may be an indication of the difference of the mechanical properties of the substance composing these craters.* (Plate 10.)

A detailed survey of the smaller ray craters would be valuable, in order to establish whether the relations between diameter, depth and height of rim are consistently different in the two types of ground. Restricting attention to ray craters will assure that the results are not confused by an unknown amount of erosion.

* Lenham (1955) has noted the consistently different relation between the size of crater and the length of its rays in the two types of ground for many other ray craters.

10. *The "tidal" bulge.*—If no large-scale melting of the Moon has to be postulated, then the arguments concerning its "fossil tidal bulge" no longer apply. The triaxial shape of the Moon appropriate to the equilibrium shape at a closer range to the Earth than the present has been used as an indication of its history; but if the Moon has always been as cold as it is now then there is no need to suppose that it has ever been of a shape satisfying the hydrostatic equation. If the long-term stiffness and strength of the material is adequate to have supported the present departure from hydrostatic equilibrium for some thousand million years, then it can be supposed that it could always stand such a departure. Only when the surface was interpreted as showing large-scale lava flows it had to be supposed that the equilibrium shape was possessed when the lava solidified, and that the present disequilibrium was an indication of the change of the Moon's orbit that had occurred since then.

If the Moon was formed by the collection of solid pieces, some of substantial size, then it must be expected to possess departures from hydrostatic equilibrium of the greatest amount that the material allows, and therefore the size of the excess tidal bulge is no more than an indication of the present material properties. Bodily tidal friction would again have to be invoked to stop the Moon's spin and oscillation relative to the Earth. It is clear that this would leave it rotating with whatever excess bulge it possessed facing the Earth. No indication of past history is then contained in the present dynamical situation other than that enough time has elapsed for the tides to act in that way. A knowledge of the internal friction of lunar material is required to estimate that time.

11. *Conclusion.*—The present views are incomplete in the sense that many processes that may be concerned require further investigation. But the principal aim has been to show that there is no real requirement for the hypothesis of lava flow and to indicate the processes that may be expected to be occurring. If some of those processes are not readily susceptible to theoretical or observational investigation, this provides no argument for leaving them out of consideration. It would be extremely difficult to investigate the details of any one of the terrestrial erosion processes in a theoretical manner if it had not been seen in fine detail; but not nearly so difficult, once the conditions on the planet are known, to suggest major agencies for the erosion process.

The views are also incomplete in the sense that there are many features that have been reported on the lunar surface for which still no explanation has been offered. But they are not the common, widespread features. There is no doubt that much fine detail must remain unexplained for much longer. One must expect that of the vast amount of fine detail that can be seen only the commonly recurring features will fit into a simple theory; that many of the rare features will reflect events of a degree of improbability that puts them well outside the useful range of speculation. There are chance effects of grazing angles of incidence of meteorites, or of tight clustering of meteorites resulting from a large piece broken up by a close passage to the Earth, which may have occurred a few times, but not often enough to establish a recognizable class of features. The apparent inexplicability of such rare markings should not be used an an observational test of a theory.

Royal Greenwich Observatory,
Hurstmonceux, Sussex.
1956 *January* 12

References

BALDWIN, R. B., *The Face of the Moon*, Univ. Chicago Press, 1949.
BUETTNER, K., *Pub. Astr. Soc. Pacific*, **64**, 11, 1952.
KUIPER, G. P., On the origin of the lunar surface features, *Proc. Nat. Ac. Sc.*, **40**, No. 12, 1954.
LENHAM, A. P., *J.B.A.A.*, **65**, 241, 1955.
PETTERSON, H., and ROTSCHI, H., *Geochim. et Cosmochim. acta*, **2**, 81, 1952.
UREY, H. C., *The Planets*, Yale Univ. Press, 1952.

THE EAST–WEST ASYMMETRY OF SOLAR FLARES AND THEIR RADIO EMISSION

J. S. Hey and V. A. Hughes

(Received 1955 July 12)

Summary

The distribution in longitude from central meridian of flares with associated radio bursts at $\lambda = 4{\cdot}1$ m during 1947–50 inclusive is analysed in relation to high geomagnetic activity. It is concluded that the asymmetry in the distribution, with greater radio burst emission from flares East of central meridian, appears to be a general flare characteristic which may be attributed to absorption and refraction in the solar atmosphere above the flare. The asymmetry is emphasized in flare regions associated with the occurrence of high geomagnetic activity; it is possible that at least part of the asymmetry, and particularly a reduction in radio emission which appears near central meridian, may be due to the early stages of corpuscular stream emission.

It is suggested that an asymmetry in the distribution of numbers of flares is similarly attributable to Hα absorption in an asymmetrical structure of the solar atmosphere above the flare.

---

1. *Introduction.*—In an analysis of the bursts of solar radio emission at $\lambda = 4{\cdot}1$ m during part of 1946–7, Hey, Parsons and Phillips (**1**) found that radio bursts appeared to be more often associated with flares on the eastern half of the visible solar disk than on the western half. It was tentatively suggested that the asymmetry might be an absorption effect in solar corpuscular streams emitted from the flare-active regions; owing to the rotation of the Sun, the envelopes of such streams are curved and any absorption effects consequently would be asymmetrical.

If the above supposition is correct it seems probable that a correlation should exist between the asymmetry and geomagnetic activity, which might be taken as an index of the arrival in the Earth's atmosphere of solar corpuscular streams. The relations between geomagnetic activity, solar flares, and radio bursts at $\lambda = 4{\cdot}1$ m are here analysed over the years 1947–50 inclusive, and possible explanations are discussed. Evidence for asymmetry in the number of flares observed visually in Hα is also considered.

2. *The observation of radio bursts at* $\lambda = 4{\cdot}1$ *m.*—The receiving aerial was a half-wave dipole with reflector at a height $\lambda/2$ above the ground. The beam was maintained on the azimuth of the Sun by automatic rotation of the receiver cabin on which the aerial was supported. Owing to the effect of ground reflection, the power received from solar radio emission varies with the elevation of the Sun. A calibrated radio receiver with stabilized power supply was used to record the output, and the radio power flux at the aerial could be determined

after making the necessary correction for the elevation of the Sun. With simple apparatus of this kind it is only possible to record radiation above a certain minimum level of intensity. For convenience in analysis, a classification in scale numbers was used as follows:

Scale number	Radio power flux in units of 10^{-21} watts $m^{-2}(c/s)^{-1}$ for horizontal polarization
0	<2
½	2–10
1	10–50
2	50–150
3	>150

The scale number 0 refers either to the absence of bursts or to intensities below 2×10^{-21} W $m^{-2}(c/s)^{-1}$, which level corresponds to about 10 times the intensity of the quiet Sun. Intensities below this level could not be measured with the apparatus, and at low elevations there was uncertainty in estimating intensities with scale number ½. The analysis was also restricted to bursts of duration longer than ¼ minute, and intensity exceeding 50 per cent of the prevailing mean level. A radio burst was arbitrarily classified as associated with a flare if it occurred within the flare period extended by 30 minutes before the first observation of the flare to 30 minutes after the last observation.

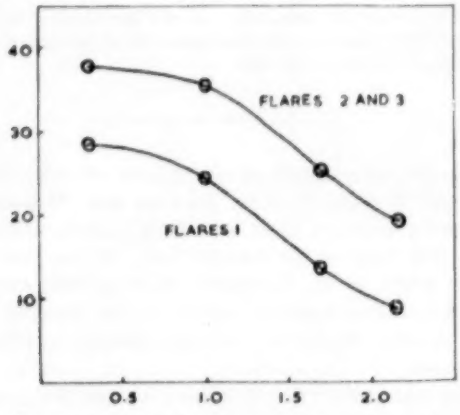


FIG. 1.—*Dependence of radio burst intensity at $\lambda = 4{\cdot}1$ m on flare importance.*
Ordinates : N = percentage of flares with radio bursts of intensity > I.
Abscissae : $\log_{10} I$, where I = radio burst intensity in units of 10^{-21} watts $m^{-2}(c/s)^{-1}$ for horizontal polarization.

The characteristics of the radio bursts at $\lambda = 4{\cdot}1$ m associated with flares have been described elsewhere (**1, 2**), where it is shown that although radio bursts may occur in the absence of flares, a large proportion are correlated with the occurrence of flares. The higher the flare importance the greater the chance of an associated radio burst. The relationship between flare importance and radio burst intensities is illustrated in Fig. 1.

The E–W asymmetry across the solar disk has been examined in terms of the number of radio bursts with intensity scale number >0, which occur in association with solar flares. The preponderance of associations between bursts

and flares on the eastern half of the Sun is demonstrated in Fig. 2 for the years 1947–50, with additional confirmatory evidence provided by observations during part of 1946 and 1951. The proportion of bursts with flares for the years 1947–50 is 32·1 per cent East as compared with 28·2 per cent West. The radio asymmetry

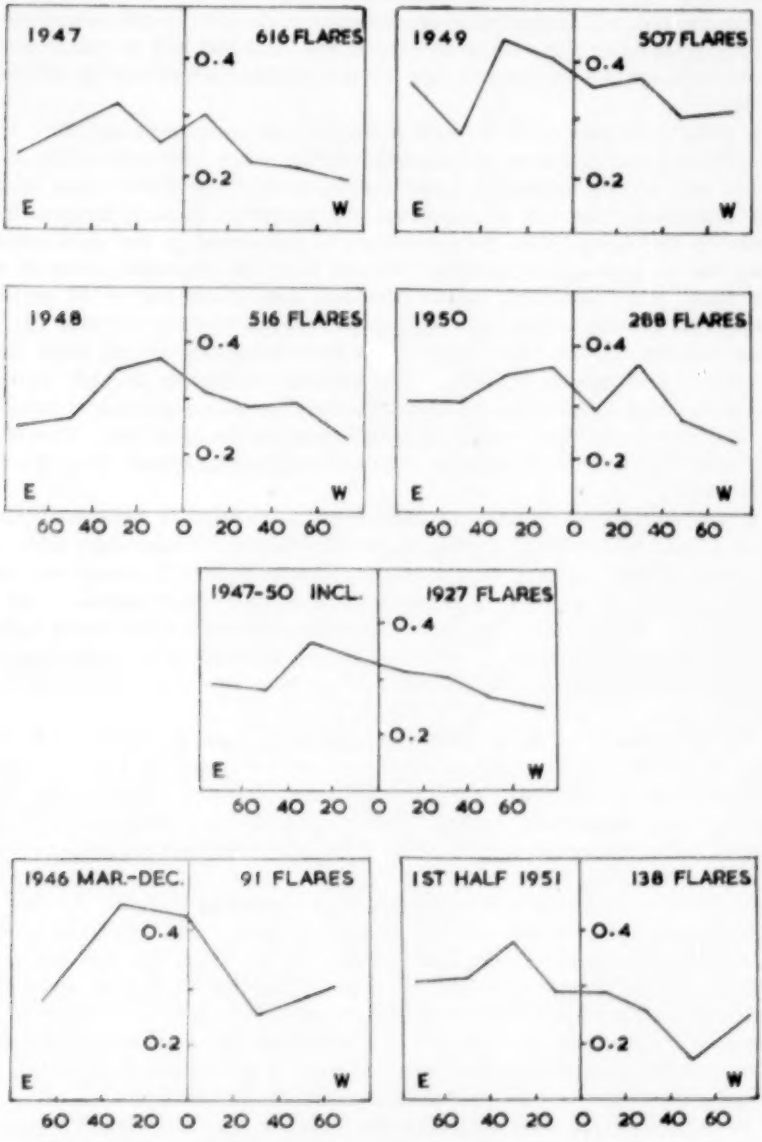


FIG. 2.—*Distribution of proportion of flares with radio bursts at* $\lambda = 4{\cdot}1$ *m.*
Ordinates: proportion of flares with radio bursts.
Abscissae: solar longitude from central meridian.

cannot be attributed to a chance excess of high-intensity flares East since the percentage of flares of importance 2 and 3 is 17·3 East as compared with 17·4 West. The statistical significance of the asymmetry has increased each year, although with a total of over 2000 flares the chance probability of an E–W asymmetry of the observed magnitude is about 5 per cent. Consequently there is difficulty in sub-dividing the observations for analysis. Nevertheless, there is a strong probability that the asymmetry is real, and this will be reinforced if relationships are found between the asymmetry and other aspects of solar activity.

3. *Relation between the E–W radio asymmetry and geomagnetic activity.*—It is generally accepted that most geomagnetic activity arises from solar corpuscular streams entering the terrestrial upper atmosphere. Even if the origin of the E–W asymmetry may not be attributed to absorption in such streams it is of interest to examine how the asymmetry is influenced by the solar regions giving rise to geomagnetic activity. Except near the minimum phase of the solar cycle, when recurrent magnetic storms may occur due to M regions, geomagnetic activity is associated with active sunspot regions. Newton (3) has shown that the regions where flares occur most frequently are the most likely to produce geomagnetic activity. The greatest probability of high activity, indicated by high values of the magnetic character figure C, occurs one or two days after the central meridian passage of these regions on the solar disk. This lag is determined by the time of travel of the main corpuscular stream from the Sun to the Earth.

It was accordingly decided to analyse the proportion of flares with radio bursts E and W respectively from eight days before to four days after the occurrence of high values of daily magnetic character figure C. About eight days preceding high values of C a preponderance of flare-active regions may be expected near the E limb; and about four days after high values these regions will have moved over to the W limb. The average value of C during 1947–50 inclusive is approximately 0·7, and high values were arbitrarily chosen as those with $C \geqslant 1{\cdot}1$.

The proportions of flares with radio bursts E and W respectively from eight days before to four days after $C \geqslant 1{\cdot}1$ are compared in Fig. 3. These diagrams show at once that a relationship exists between the East and West radio emission from flares, with the eastern radio effects predominating on the average. The maximum eastern radio burst activity occurs between two and eight days preceding high values of C. This suggests that the eastern regions where the corpuscular radiation may then be supposed to originate, produce an enhanced proportion of radio bursts. As the Sun rotates, this radio burst activity subsequently diminishes on the East and rises on the West, but does not attain as high a level as it did on the East. It is noticeable that when these active flare regions are on the eastern half of the disk the radio burst activity of flares on the West is at a minimum. It is not possible to explain this in terms of absorption in corpuscular streams from the eastern region. The impression is gained that the radio asymmetry is a general characteristic of all flare regions.

As Newton (3) has shown, geomagnetic activity can be most clearly associated with regions of exceptional flare activity, especially those with high-intensity flares; and as Fig. 1 indicates that the higher the flare importance the greater the chance of a radio burst, it is understandable that the regions which are most

active in producing high-velocity streams which reach the Earth also predominate in radio bursts. However, the general trend shown in Fig. 3 is similar when flares of importance 2 and 3 are excluded, so that the correlation of high geomagnetic activity with high radio burst activity is not merely a function of flare importance. The selection of flare regions producing geomagnetic activity makes it possible to trace their radio effects across the disk as the Sun rotates. Their progress will now be determined in more detail by examining the distribution in solar longitude.

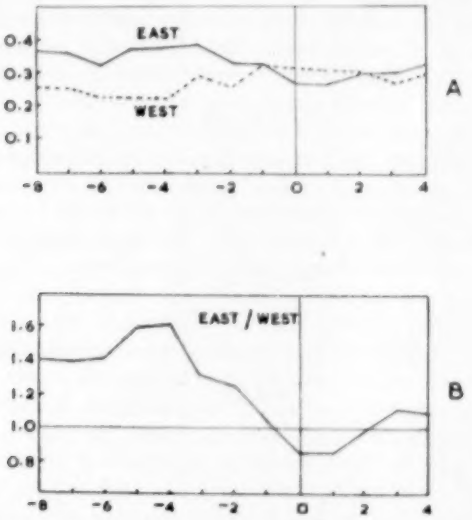


FIG. 3.—*Proportion of flares with radio bursts, 1947–50, λ = 4·1 m, on days before and after* $C \geqslant 1{\cdot}1$.

Ordinates : Fig. 3 A: proportion of flares with radio bursts.
Fig. 3 B: ratio of proportions, East/West.
Abscissae : days before (−) or after (+) occurrence of $C \geqslant 1{\cdot}1$.

4. *The relation between geomagnetic activity and the E–W distribution of flares and associated radio bursts.*—When the distribution of the proportion of flares with radio bursts is plotted across the Sun in relation to $C \geqslant 1{\cdot}1$, as shown in Fig. 4, it is surprising to find that, instead of a peak in radio burst activity moving steadily across the Sun from East to West, the distribution retains some similarity in form throughout. There is a marked peak near 20° E which is particularly enhanced a few days before magnetic activity is high. A less well-marked peak appears on the W side following high magnetic activity. There is, however, no occurrence of a peak near central meridian. It is thus evident that the radio burst activity of these regions associated with high values of C varies in such a way that the greatest activity is produced in the East, then there is a reduction of radio burst activity near central meridian followed by a partial recovery of activity West. The features of similarity in all the distributions of Fig. 4 support the idea that flares generally tend to behave in this way. If the distributions are each divided by the mean for eight days before to four days after $C \geqslant 1{\cdot}1$ it becomes at once apparent that the distributions can be represented as the product of the

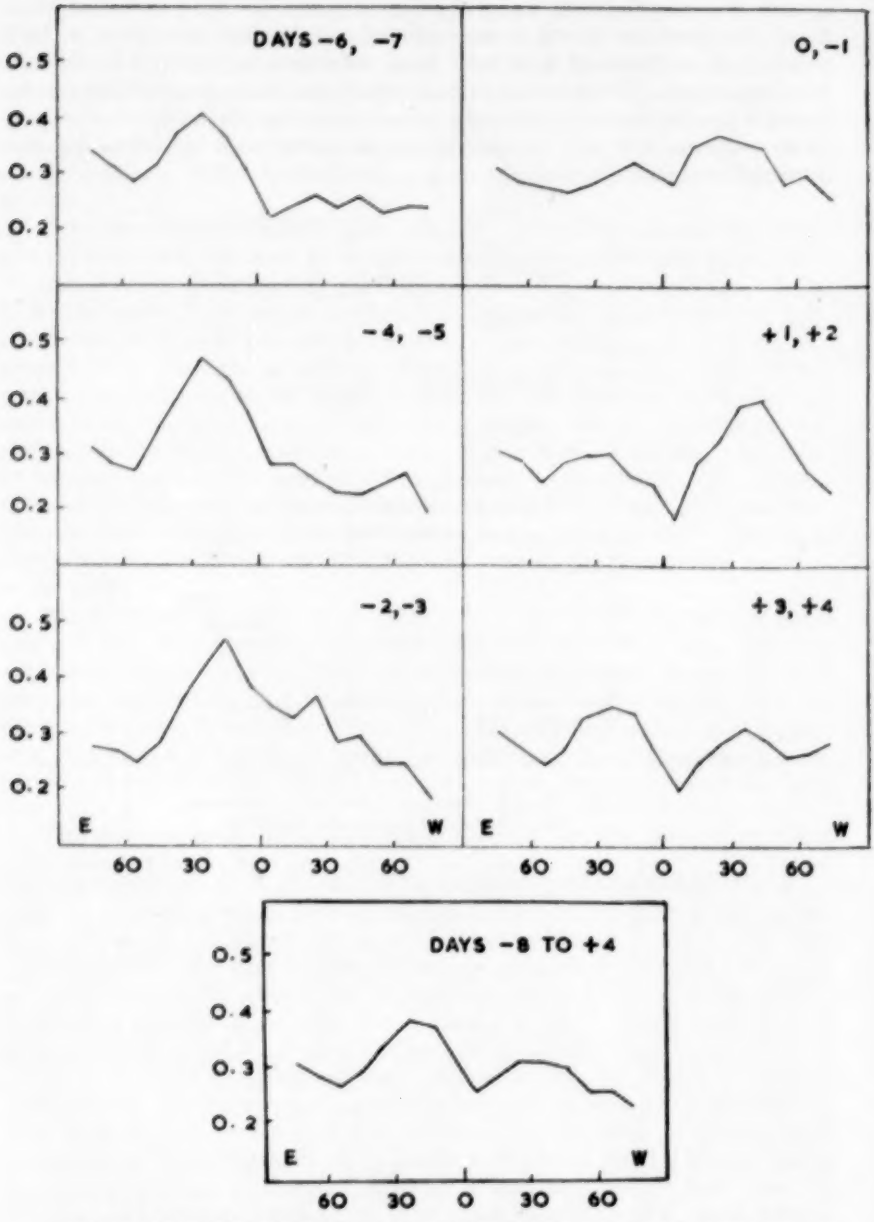


FIG. 4.—*Distributions of proportions of flares with radio bursts, 1947–50, λ = 4.1 m on days before (−) and after (+) occurrence of C ≥ 1·1.*
(Averages over 30° intervals are plotted each 10°.)
Ordinates : proportions of flares with radio bursts.
Abscissae : solar longitude from central meridian.

mean asymmetrical distribution and a wave of high radio burst activity moving across the Sun. The selection of flares associated with high values of C, by increasing the proportions of bursts at certain longitudes, helps to reveal more clearly the influence of the asymmetrical distribution. The most notable difference in the mean distribution associated with high values of C, as compared with that for all flares irrespective of C, is the marked dip near central meridian.

5. *Discussion of the radio asymmetry.*—The results so far have suggested that the asymmetry is a general characteristic of the radio bursts emitted by flares, and the most likely explanation appears to be that it is caused by absorption and refraction in the solar atmospheric structure above the flare, including corpuscular streams emitted from the region. It is also possible that in flares which are not followed by increased values of C streams may occur but with insufficient velocity to escape. If absorption in corpuscular streams does affect the radio emission it must in any case occur near the flare if it is assumed that the stream expands. For the radio absorption coefficient is proportional to N^2 where N is the electron density. If the corpuscular stream expands, say in a cone, then N is proportional to r^{-2} where r is the distance from the flare, and the absorption coefficient is thus proportional to r^{-4}. The consequence of this is that the main radio absorption occurs near the origin of the stream where the electron density is high. If the electron density is sufficiently high for the refractive index to decrease appreciably, the absorption is even greater, since the absorption coefficient is inversely proportional to the refractive index.

Whatever the cause of the asymmetry, the effect can be described in terms of an angular distribution of radiation from the flare. There is a zone of maximum emission between 10° and 40° westward from the normal, corresponding to the peak observed when the flare regions are between 10° and 40° East of central meridian. If this zone is due to a deficiency of electrons with high electron density on each side, the latter would reinforce the radio emission inside the zone by refraction as well as reducing the radio emission outside by absorption and refraction.

It must of course not be overlooked that the results are the average of a large number of flares from sunspot regions with heterogeneous individual characteristics. An average asymmetry in the structure of the solar atmosphere above these regions appears reasonable since sunspot groups whether bipolar or unipolar are themselves most often asymmetrical East–West formations.

The order of "optical depth" τ that would be needed to account for the observed asymmetry if it were assumed to be attributed to absorption may be determined. The ratio of intensities East (I_E) and West (I_W) is

$$I_E/I_W = \exp(-\tau)$$

or

$$\tau = 2{\cdot}3\,[\log_{10} I_E - \log_{10} I_W].$$

The distribution of radio intensities East and West is shown in Fig. 5. Considering only the part of the graphs corresponding to scale numbers $\geqslant \frac{1}{2}$, since the accuracy of the graph at lower intensities is open to doubt, it is seen that the intensities West are lower than those East by approximately $\log_{10} I = 0{\cdot}2$. This represents on the average a greater optical depth West by $\tau \approx 0{\cdot}5$. In the quiet Sun a ray at 4 m wave-length can attain a greater optical depth than this in the corona. The optical depth increases at lower heights as the electron

density increases, and in fact normally a ray cannot penetrate to the chromosphere. There does not appear to be any difficulty in supposing a difference in optical depths of the order of 0·5 may occur due to asymmetry in the solar atmospheric structure above the flare. It is tempting to assume that part at least of the attenuation is associated with the early stages of ejection of an ionized stream from the region. The reduction in radio bursts when flare-active regions are near central meridian lends support to this.

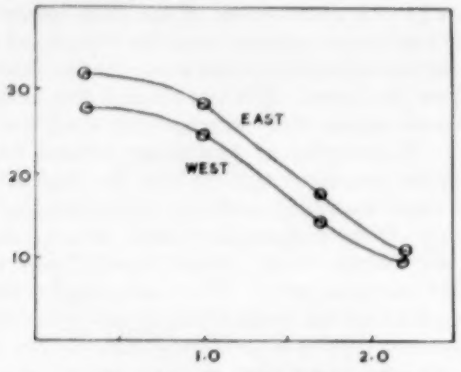


FIG. 5.—*Intensities of radio bursts East and West.*
Ordinates : N = percentage of flares with radio bursts of intensity > I.
Abscissae : $\log_{10} I$, where I = radio burst intensity in units of 10^{-21} watts $m^{-2}(c/s)^{-1}$ for horizontal polarization.

6. *East–West distribution of flares in Hα.*—Flares are normally observed in Hα light. During the investigation of the 4 m radio emissions it was noticed that often the distribution of flares seemed to have some resemblance to the radio distribution. When the distribution of all the flares here analysed in the years 1947–50 was plotted, an East–West asymmetry was noticed as shown in Fig. 6(*a*). As a certain amount of selection of data is involved (for example, if there is a strong sunspot emission which may mask a radio burst, then the flare is omitted from analysis unless a burst occurs which is clearly distinguishable), it was decided to test the East–West distribution for all flares listed in the I.A.U. *Quarterly Bulletin on Solar Activity* for the years 1947–50. These observations are plotted in Fig. 6(*b*), which again shows the asymmetry. Link and Kleczek (**4**) and Behr and Siedentopf (**5**) had previously reported such an asymmetry in their analyses of flares from 1935–47 and 1935–50 respectively, so that our result was entirely in accordance with their conclusions.

An analysis may be made to find, as in the radio bursts, how the asymmetry in the distribution of flares is influenced by its relation to high geomagnetic activity. As discussed earlier, the magnetic character figure tends to be high following central meridian passage of flare-active regions. Fig. 7 shows the percentage of occurrences of values of $C \geqslant 1{\cdot}1$ with respect to the day of central meridian passage of regions contributing flares to the analysis, each region being weighted by the number of flares. Conversely, this represents the expected distribution on the solar disk of solar flares which occur on a day when $C \geqslant 1{\cdot}1$, assuming flares equally probable at all solar longitudes. Each day corresponds to a rotation of 13°·3 in solar longitude. Thus N days before $C \geqslant 1{\cdot}1$ the expected distribution must be displaced to the East by N days $= 13^{\circ}{\cdot}3N$.

The reduction of flare visibility towards the solar limb must however be taken into account. Waldmeier (**6**) and Behr and Siedentopf (**5**) indicate that a factor $a + b\cos\theta$, where θ is angular distance from the centre and a, b are suitably chosen constants, is approximately correct. Since the results are here being analysed in order to determine departures from the average by choosing flares in relation to high values of C, the observed distribution for all the flares, as shown in Fig. 6(*a*), must be used as the visibility factor.

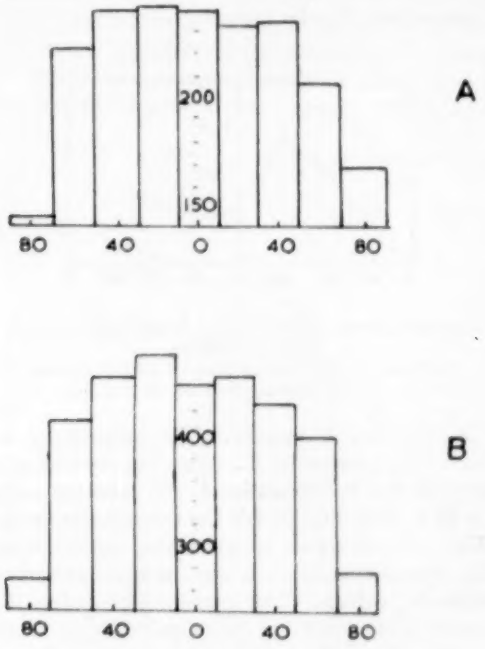


FIG. 6.—*Distribution of flares in Hα, 1947–50.*
(a) flares included in λ=4·1 m analysis. (b) all flares.
Ordinates: number of flares.
Abscissae : solar longitude from central meridian.

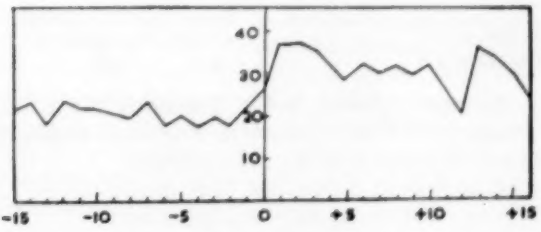


FIG. 7.—*Percentage of occurrences of $C \geqslant 1{\cdot}1$ before and after flare-active regions pass central meridian, 1947–50.*
Ordinates : percentage of occurrence of $C \geqslant 1{\cdot}1$.
Abscissae: day relative to C. M. P. of flare region.

The expected distributions of flares have been calculated by this method from eight days before to four days following the occurrence of $C \geqslant 1{\cdot}1$, and these have been compared with observed distributions. A similarity between the radio distributions and the flare distributions was found to be evident, and this is illustrated by Fig. 8. This indicates that the East–West asymmetry is emphasized by selecting active regions responsible for high magnetic activity. A notable feature is also the reduction in the observed compared with the expected numbers of flares near central meridian, which is similar to the observed reduction in the proportion of radio bursts.

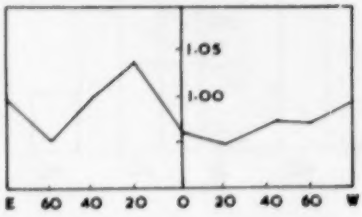


FIG. 8.—*Ratio of observed/expected numbers of flares, 8 days before to 4 days after occurrence of* $C \geqslant 1{\cdot}1$, 1947–50.
Ordinates : ratio, observed/expected numbers of flares
Abscissae : solar longitude from central meridian.

7. *Discussion of Hα flare asymmetry.*—The asymmetry in flares strongly suggests that a similar explanation to that offered in the case of radio bursts may apply, namely that it is due to attenuation in the solar atmospheric structure in the vicinity of the flare, including in this the corpuscular streams emitted from the region. Unlike radio at metre wave-lengths, optical refraction cannot be important and the attenuation will be due to hydrogen absorption and will occur in chromospheric regions. The average attenuation West as compared with East may be estimated by plotting the intensity distributions. The intensity distributions may conveniently be derived from the lists of completely observed flares quoted by Link (**7**) for the more extensive period 1936 to 1950. As Ellison (**8**) has shown, the Hα line widths may be taken as a measure of flare intensities and may be approximately classified according to flare importance as follows:

Flare class	1	2	3
Hα line width in A	2–4	4–6	6

In this way, the graphs plotted in Fig. 9 showing the E and W relations between flare frequency and the line width have been obtained. The mean East–West difference in $\log_{10} I$ is of the order of 0·02.

Hence

$$\tau = 2{\cdot}3\,[\log_{10} I_{\mathrm{E}} - \log_{10} I_{\mathrm{W}}] \simeq 0{\cdot}05.$$

This comparatively small differential absorption, of the order of 5 per cent, which is revealed by statistics may well pass unnoticed in normal individual flare observations, especially if it is at least partly due to stationary features of chromospheric structure. An absorption band on the blue wing of the Hα

emission lines has been observed by Ellison (9) and others. This is thought to be due to absorption in a corpuscular stream of hydrogen in the initial stages of expansion from the flare. It seems possible that reduction of observed flare numbers near central meridian associated with regions causing high geomagnetic activity may be connected with corpuscular stream absorption.

8. *Conclusions.*—The above analysis has provided evidence that the intensities of both radio and Hα radiation from flares are greater on the eastern than on the western half of the solar disk. It is suggested that the radio asymmetry is caused by absorption and refraction in an asymmetrical E–W structure in the electron distribution in the solar atmosphere above the flare. The visual flare asymmetry may similarly be caused by hydrogen absorption closer to the flare. Assuming that absorption is responsible, the average greater optical depth West compared

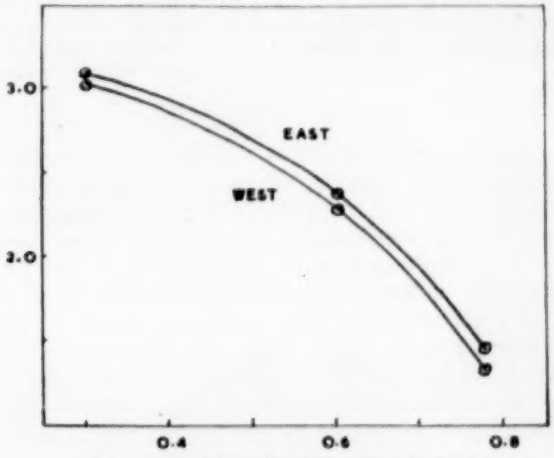


FIG. 9.—*Hα line widths of flares East and West.*
Ordinates : $\log_{10} N$ where N = number of flares.
Abscissae : $\log_{10} L$ where L = Hα line width in A.

with East is of the order of 0·5 at the radio wave-length of $\lambda = 4{\cdot}1$ m and 0·05 in Hα. It is believed that at least part of the absorption effect, particularly the reduction near central meridian which is emphasized in active regions responsible for high geomagnetic activity, may be associated with absorption in the initial stages of the expulsion of solar corpuscular streams. The correlation found between high radio burst activity and high geomagnetic activity indicates a connection between radio and corpuscular emission from flares.

Finally, it must be emphasized that the conclusions are based mainly on the analysis of data from the years 1947–50 inclusive. Although there appears strong evidence for accepting their significance, it is most desirable that more radio data should be analysed as they become available so as to establish with certainty the observed asymmetries as general solar characteristics.

Ministry of Supply,
Radar Research Establishment,
Leigh Sinton Road,
Malvern, Worcs.:
1955 *July* 10.

References

(1) J. S. Hey, S. J. Parsons and J. W. Phillips, *M.N.*, **108**, 354, 1948.
(2) J. S. Hey, *M.N.*, **113**, 653, 1953.
(3) H. W. Newton, *M.N., Geophys. Suppl.*, **5**, 321, 1949.
(4) F. Link and J. Kleczek, *Comptes Rendus*, **227**, 827, 1948.
(5) A. Behr and H. Siedentopf, *Zs. f. Ap.*, **30**, 177, 1952.
(6) M. Waldmeier, *Astr. Mitt. Eidgen. Sternw. Zürich*, No. 153, 1948.
(7) F. Link, 8th Report of the Commission for the Study of Solar and Terrestrial Relationships, 35, 1954.
(8) M. A. Ellison, 8th Report of the Commission for the Study of Solar and Terrestrial Relationships, 27, 1954.
(9) M. A. Ellison, *M.N.*, **103**, 3, 1943.

THE Hα EMISSION OF PROMINENCES

J. T. Jefferies

(Communicated by R. G. Giovanelli)

(Received 1955 August 18)*

Summary

It is shown that Hα radiation from the solar disk is strongly reflected by prominences and is an important component of their radiation.

An analysis of their observed Hα emission at the limb indicates a range of kinetic temperatures of about 10^4 to 2×10^4 deg. K and electron concentrations of about 10^{10} to 5×10^{10} cm^{-3} for stable prominences.

Two alternative explanations are given for the double peaked profiles of Hα sometimes observed in solar prominences, depending on whether scattering is coherent or incoherent.

---

1. *Introduction.*—Measurements of the central intensities and half-widths of the Hα radiation emitted by stable prominences on the Sun's limb have been made by Conway (**1**) and by Ellison (**2**). In a later publication Conway (**3**) has interpreted some of her profile observations in terms of a theoretical expression for the emission from a model prominence and obtained kinetic temperatures in the range $1{\cdot}1 \times 10^4$ to $1{\cdot}8 \times 10^4$ deg. K. The absolute intensity measurements, however, require for their interpretation a knowledge of the excited state populations of hydrogen as a function of temperature and density, such as have been discussed recently by Jefferies and Giovanelli (**4**). On the basis of results obtained in (**4**) we present here an analysis of the observational results of (**1**) and (**2**) and deduce some of the physical properties of prominences consistent with the observational material.

To analyse the data given by Conway (**1**) and Ellison (**2**), we shall compute the Hα emission from a model hydrogen prominence in the form of a slab of uniform kinetic temperature, electron concentration and thickness. The slab is supposed to be perpendicular to the Sun's surface, oriented so that its minimum thickness is in the line of sight and illuminated by chromospheric and photospheric radiation incident from a hemisphere the diametrical plane of which is at right angles to the plane of the prominence. The intensity of the Hα radiation incident on the prominence may be related to that at the centre of the disk by a limb-darkening law $I_1(\mu') = \frac{2}{5} I_1(1)[1 + \frac{3}{2}\mu']$, where $I_1(\mu')$ is the intensity at an angle $\cos^{-1}\mu'$ to the normal to the Sun's surface. The present results are rather insensitive to departures from this law. The radiation from this model prominence will consist of two parts, one the diffusely reflected radiation and the other the emission of the material itself. These contributions depend on the degree of coherence in frequency of the scattered radiation. We shall consider here two cases such that the scattering is either completely or partially coherent; in each case we shall obtain expressions for the self-emitted and diffusely reflected radiation and compare the computed emission profiles with observational results.

* Received in original form 1955 March 29.

2. *The equation of transfer.*—The monochromatic specific intensity $I_\nu(\tau,\mu)$ of the radiation in a plane parallel slab such as our model prominence is given by the equation of transfer

$$\mu \frac{dI_\nu(\tau,\mu)}{d\tau_\nu} = I_\nu(\tau,\mu) - \mathfrak{S}_\nu(\tau,\mu), \tag{1}$$

where $\cos^{-1}\mu$ is the angle the radiation makes with the normal to the slab, τ_ν is the optical thickness and $\mathfrak{S}_\nu \equiv j_\nu/\kappa_\nu$ is the source function, j_ν and κ_ν being respectively the monochromatic emission and absorption coefficients. For a medium which scatters coherently and isotropically, the source function may be written

$$\mathfrak{S}_\nu(\tau) = \epsilon_\nu/\kappa_\nu + (1-\lambda)J_\nu/4\pi \tag{2}$$

where ϵ_ν is the true emission coefficient, J_ν is the total monochromatic intensity $\equiv \int_{4\pi} I_\nu\, d\omega$ and λ is a scattering parameter defined so that $(1-\lambda)$ is the fraction of absorbed radiation which is subsequently scattered.

The specific intensity of the radiation emerging from the atmosphere follows on solution of (1) and is given, neglecting frequency subscripts, by

$$I(0,\mu) = I(\tau_1,\mu)\exp(-\tau_1/\mu) + \int_0^{\tau_1} \mathfrak{S}(t,\mu)\exp(-t/\mu)\, dt/\mu, \tag{3}$$

τ_1 being the total optical thickness of the atmosphere.

To evaluate I, we substitute in (2) an approximate value of J obtained as the solution of Eddington's approximation to the equation of transfer,

$$\frac{1}{3}\frac{d^2J}{d\tau^2} = \lambda J - 4\pi\frac{\epsilon}{\kappa}. \tag{4}$$

If ϵ/κ and λ are independent of τ, the solution of (4) is

$$J = \frac{4\pi\epsilon}{\kappa\lambda} + A\exp(\sqrt{3\lambda}\tau) + B\exp(-\sqrt{3\lambda}\tau) \tag{5}$$

A and B being integration constants. Since the total intensities on each face of the slab are equal, it follows that $B = A\exp(\sqrt{3\lambda}\tau_1)$.

The second boundary condition follows readily from the definitions of the total intensity J, given above, and of the net flux $F \equiv \int I\mu\, d\omega$. Assuming that, at the surface of the slab, the specific intensity of the outwardly directed radiation is independent of μ and that the inwardly directed radiation, coming from the solar disk, follows a law of darkening of the form given above, it follows readily that

$$J - 2\pi I' = 2F \tag{6}$$

where $I' = 0{\cdot}678 I_1(1)$.

Using the Eddington approximation $F = \frac{1}{3}\, dJ/d\tau$, the integration constants A and B are easily found and the total intensity takes the form

$$J = \frac{4\pi\epsilon}{\kappa\lambda}\left[1 - \frac{\exp(-\sqrt{3\lambda}\tau) + \exp[\sqrt{3\lambda}(\tau-\tau_1)]}{D}\right] + \frac{2\pi I'}{D}\left[\exp[\sqrt{3\lambda}(\tau-\tau_1)] + \exp(-\sqrt{3\lambda}\tau)\right] \tag{7}$$

where $D = 1 + 2\sqrt{\lambda/3} + (1 - 2\sqrt{\lambda/3})\exp(-\sqrt{3\lambda}\tau_1)$.

Substituting (7) into (2) and carrying out the integration in (3) it follows that the specific intensity of the emergent radiation in the direction $\mu = 1$, i.e. at right angles to the plane of the prominence, is

$$I(0, 1) = \frac{\epsilon}{\kappa\lambda}\left[1 - \exp(-\tau_1) - F(\beta, \tau_1)\right] + \frac{I'}{2} F(\beta, \tau_1) \tag{8}$$

where $\beta = \sqrt{3\lambda}$ and

$$F(\beta, \tau_1) = \frac{1-\lambda}{D}\left\{\exp(-\beta\tau_1)\left[\frac{\exp[(\beta-1)\tau_1]-1}{\beta-1}\right] + \frac{1-\exp[-(\beta+1)\tau_1]}{\beta+1}\right\}.$$

The term $I(\tau_1, \mu)\exp(-\tau_1/\mu)$ appearing in (3) represents the radiation transmitted directly (i.e. not diffusely) through the prominence. For a prominence projecting beyond the limb this term is zero since there is no incident radiation in the line of sight.

We shall be interested in the ratio R_ν of the intensity $I_\nu(0, 1)$ to the intensity I_w of the continuous spectrum at the centre of the disk just outside the Hα line. This ratio may be written

$$R_\nu = \frac{\epsilon/\kappa\lambda}{I_w}\left[1 - \exp(-\tau_1) - F(\beta, \tau_1)\right] + 0{\cdot}339\,r_\nu F(\beta, \tau_1) \tag{9}$$

where r_ν represents the contour of the Hα absorption line at the centre of the disk. The two terms on the right of (9) represent respectively the emission from the prominence itself and the diffusely reflected incident radiation. The function $F(\beta, \tau_1)$ differs by less than 5 per cent from the accurate expression obtained by Chandrasekhar (5) for the diffuse reflection and transmission of such a slab, at least over the range available for comparison, i.e. where Chandrasekhar's X and Y functions have been tabulated. The function $F(\beta, \tau_1)$ for $\lambda = 6 \times 10^{-3}$ is shown in Fig. 1.

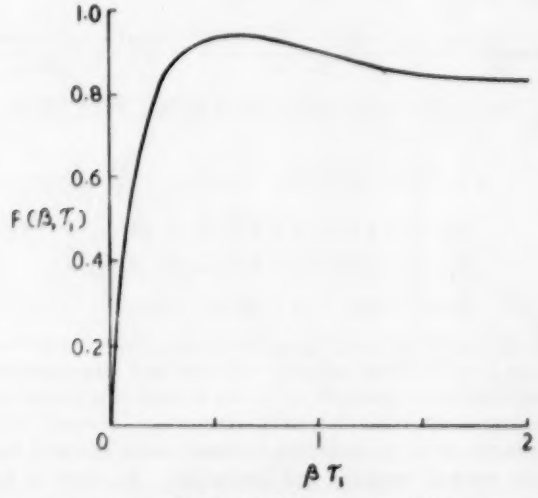


FIG. 1.—*The function* $F(\beta, \tau_1)$ *for* $\lambda = 6 \times 10^{-3}$.

2.1. *Mixed coherent and non-coherent scattering.*—According to Woolley and Stibbs (**6**) the redistribution of frequency on scattering results in a source function of the approximate form

$$\mathfrak{S}_\nu = \frac{1-\lambda}{4\pi}[aJ_\nu + bJ_{\nu_0}] + \epsilon_\nu/\kappa_\nu \qquad (10)$$

where a and b are defined by

$$a = \frac{\delta_k}{\delta_j + \delta_k}, \qquad b = \frac{\delta_j}{\delta_j + \delta_k},$$

k and j referring respectively to the upper and lower states of the atom, and the quantities δ_j and δ_k being defined by

$$\delta_j = \sum_{i<j} A_{ji}/4\pi, \qquad \delta_k = \sum_{i<k} A_{ki}/4\pi,$$

the A's being spontaneous transition rates.

The specific intensity $I(0,\mu)$ of the radiation emerging at an angle $\cos^{-1}\mu$ to the normal to the prominence is given by (3) where $\mathfrak{S}_\nu$ is now given by (10). Expressions for the total intensities J_ν and J_{ν_0} may be obtained from the approximate equation, analogous to (4)

$$\frac{1}{3}\frac{d^2J}{d\tau^2} = J - (1-\lambda)[aJ + bJ_0] - 4\pi\epsilon/\kappa \qquad (11)$$

where J_0 represents the central intensity; remaining frequency subscripts have been omitted. The solution of (11) with the boundary conditions used above follows readily. Since, at the line centre, (11) is equivalent to (4), the solution for J_0 is the same as for the case of coherent scattering.

Substituting values of J and J_0 into (3) we find

$$I(0,1) = \frac{\epsilon}{\kappa\lambda}[1-\exp(-\tau_1) - \phi(\beta,\tau_1)] + \frac{I_0'}{2}\phi(\beta,\tau_1) + \left[\left(\frac{I'}{2} - \frac{I_0'}{2}x\frac{D_2}{D_1}\right) - \frac{\epsilon}{\kappa\lambda}\left(1 - x\frac{D_2}{D_1}\right)\right]\psi(q,\tau_1) \qquad (12)$$

where

$$\phi(\beta,\tau_1) = \frac{(1-\lambda)(b+ax)}{D_1}\left\{\frac{1-\exp[-(\beta k+1)\tau_1]}{\beta k+1} + \frac{\exp(-\tau_1)-\exp(-\beta k\tau_1)}{\beta k-1}\right\},$$

$$\psi(q,\tau_1) = \frac{a(1-\lambda)}{D_2}\left\{\frac{1-\exp[-(q+1)\tau_1]}{q+1} + \frac{\exp(-\tau_1)-\exp(-q\tau_1)}{q-1}\right\},$$

$$k = \tau_0/\tau, \qquad q = \sqrt{3[1-a(1-\lambda)]}, \qquad x = \frac{b(1-\lambda)}{b(1-\lambda)-\lambda(k^2-1)},$$

and

$$D_1 = (1+\tfrac{2}{3}\beta) + (1-\tfrac{2}{3}\beta)\exp(-\beta k\tau_1),$$
$$D_2 = (1+\tfrac{2}{3}\beta k) + (1-\tfrac{2}{3}\beta k)\exp(-\beta k\tau_1),$$
$$D_3 = (1+\tfrac{2}{3}q) + (1-\tfrac{2}{3}q)\exp(-q\tau_1).$$

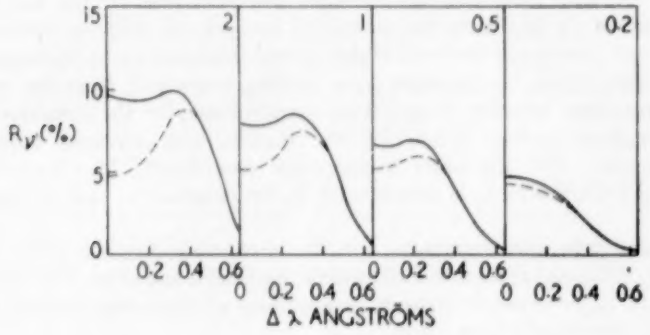
In this equation the terms multiplying $\epsilon/\kappa\lambda$ represent the self-emitted component, those multiplying I' and I_0' the diffusely reflected and transmitted component.

The diffusely reflected radiation is by no means negligible across the line for either coherent or partially coherent scattering; indeed, it predominates over the self-emission of the prominence in many cases, and as a result considerably modifies the central intensity and half-width. In view of this it may be necessary to reconsider the interpretation of other prominence spectral lines taking into account the effects of diffuse reflection of chromospheric radiation.

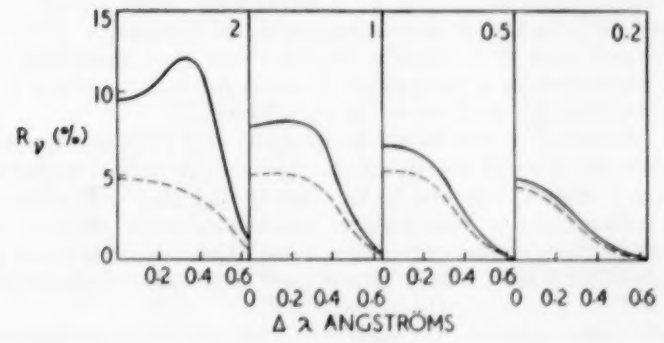
3. *The Hα intensity and contour.*—The central intensity and contour of the Hα radiation emitted by a model prominence of given τ_1^0—the optical depth at the line centre—may be computed, depending on the type of scattering considered, from either (9) or (12) if we know the values of the parameters and their frequency dependence. Expressions for the quantities ϵ/κ and λ are given in (4), equations (7.5) and (7.6); r_ν is taken from Fig. 5 of the author's paper (7) and is the mean of four sets of observations of the Hα contour at the centre of the disk. The value of λ is almost constant for the various temperatures and electron concentrations used here and for simplicity is taken to be constant and equal to 6×10^{-3}. The frequency dependence of τ_1 is taken as the standard Doppler expression

$$\tau_1 = \tau_1^0 \exp\left[-\left(\frac{c\Delta\nu}{v\nu}\right)^2 \right]$$

where the mean atomic velocity $v = \sqrt{2kT/M}$. Typical computed profiles, showing the magnitudes of the reflected components, are given in Figs. 2(*a*) and 2(*b*). Table I gives the self-emitted and diffusely reflected components of the intensity at the centre of Hα for a series of values of τ_1^0 for each of the temperatures and electron concentrations considered. The profiles of the



(*a*)



(*b*)

FIGS. 2 (*a*) *and* 2 (*b*).—*Hα contours of model prominences for values of* $\beta\tau_1^0$ *shown in the top right of each diagram. The diffusely scattered component is shown as a broken curve. The figure relates to a model with* $T = 10^4$ *deg. K,* $N_e = 10^{10}$ *cm*$^{-3}$ *and with low optical thickness in the Lyman continuum. Fig.* 2 (*a*) *is for the coherent and* 2 (*b*) *for the partially coherent case.*

emergent radiation may be obtained from these values using the above frequency dependence of τ_1^0 to find τ_1.

TABLE I

*Relative Contributions of the Self Emitted and Diffusely Reflected Radiation to the Hα Intensity**

		$T=10^4$deg. K				$T=1{\cdot}5\times10^4$deg. K	
$(\beta\tau^0{}_1)_{H\alpha}$	Diffuse Reflection	$N_e=10^{10}$ cm^{-3}		$N_e=5\times10^{10}$ cm^{-3}		$N_e=10^{10}$cm^{-3}	$N_e=5\times10^{10}$cm^{-3}
		$(\beta\tau)_{Le}\ll1$	$(\beta\tau)_{Le}\gg1$	$(\beta\tau)_{Le}\ll1$	$(\beta\tau)_{Le}\gg1$	$(\beta\tau)_{Le}\ll1$	$(\beta\tau)_{Le}\gg1$
4	4·57	6·14	12·1	10·2	15·5	7·25	17·2
2	4·79	4·85	9·61	8·14	12·3	5·75	13·6
0·5	5·25	1·40	2·78	2·35	3·54	1·66	3·94
0·1	2·93	0·19	0·38	0·32	0·49	0·23	0·53
0·01	0·40	0·01	0·03	0·03	0·04	0·02	0·04

* The self-emitted component is given in columns 3 to 8; the diffusely reflected component shown in column 2 is constant owing to the assumed constancy of λ. Results are given for each of the two limiting cases $(\beta\tau)_{Le}\ll1$ and $(\beta\tau)_{Le}\gg1$ for $T=10^4$ deg. K. The units of intensity are such that the continuum near Hα at the centre of the solar disk equals 100.

As shown in (4), ϵ/κ depends on the opacity of the slab to Lyman continuous radiation. As a general treatment taking this into account seems unnecessarily complex, we have computed the radiation characteristics for the two limiting cases obtained by assuming the models to be optically thick or optically thin in L_c. From known expressions for this optical thickness we can estimate where the transition region lies between these limiting cases and, from the results in the limiting cases, estimate the radiation characteristics for the transition region. This complication only arises for the models with electron temperature $T=10^4$ deg. K. For the other temperature considered, $T=1{\cdot}5\times10^4$ deg. K. the optical thickness in L_c is always small in the range of Hα optical thicknesses used.

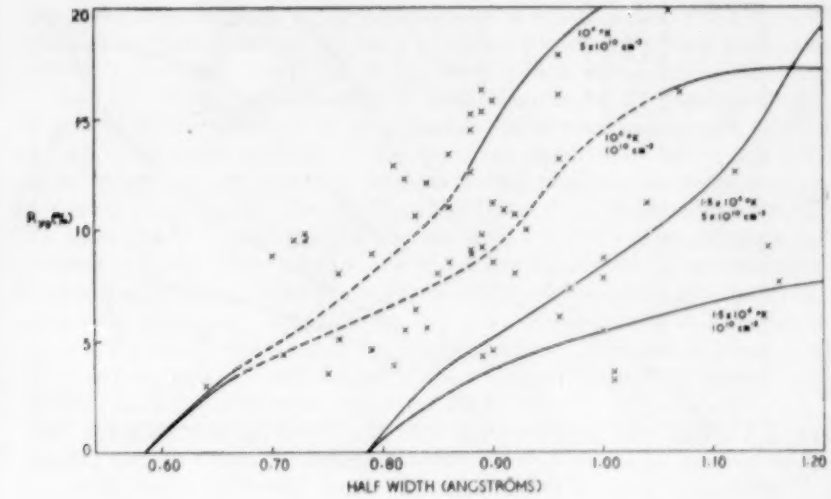
4. *Comparison with observation.*—As the observational results of Conway (1) and those of Ellison (2) are rather different in emphasis, and as we wish to consider some of the latter in detail, separate comparisons of these observations with the computed curves have been made.

Figs. 3(*a*) and 3(*b*) show for coherent and partially coherent scattering respectively a comparison of Conway's observational points with the computed curves of half-width against central intensity found by taking a series of values of τ_1^0 for each value of T and N_e. On the whole, this comparison indicates, for stable prominences, a temperature of about 10^4 to $1{\cdot}5\times10^4$ deg. K and an electron concentration of about 10^{10} to 5×10^{10} cm^{-3}.

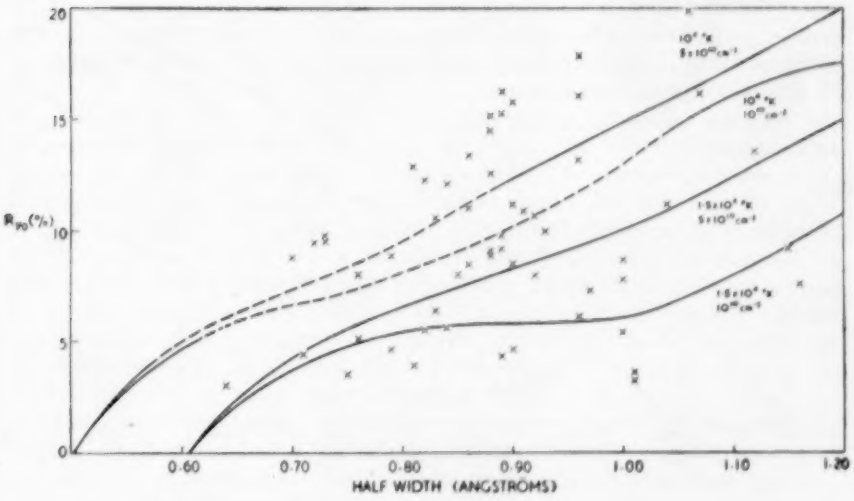
The information is insufficient to establish very definitely the values of temperature and electron concentration. There is an indication that a model with lower T and N_e may also fit the observations quite well, while, for the partially coherent case at least, a higher temperature model would be required to fit some of the lower intensity observations. Considerations of the physical thicknesses of the models, to be made later, will help to place limits on the values of N_e and T.

Ellison's measurements were made on a small number of prominences; for each one he recorded observations at a number of heights above the solar limb. These results are compared with the theoretical curves in Figs. 4(*a*) and 4(*b*). The notation used there is the same as in (2), but points corresponding to double-peaked or flat profiles are shown with a subscript F for future reference.

The extent to which any one set of observational points follows the trend of the computed curves gives an idea of the uniformity of N_e and T at various heights in a prominence. There is some indication of such a uniformity for some prominences, but much more extensive observations would be required to confirm this.



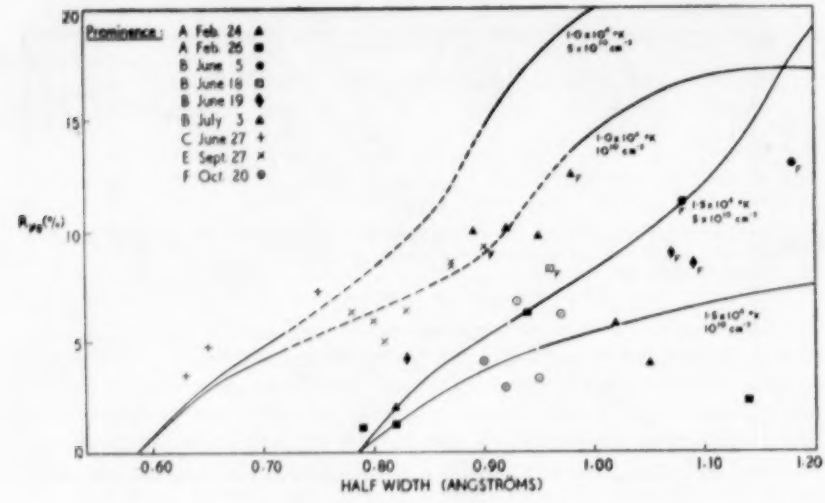
(*a*)



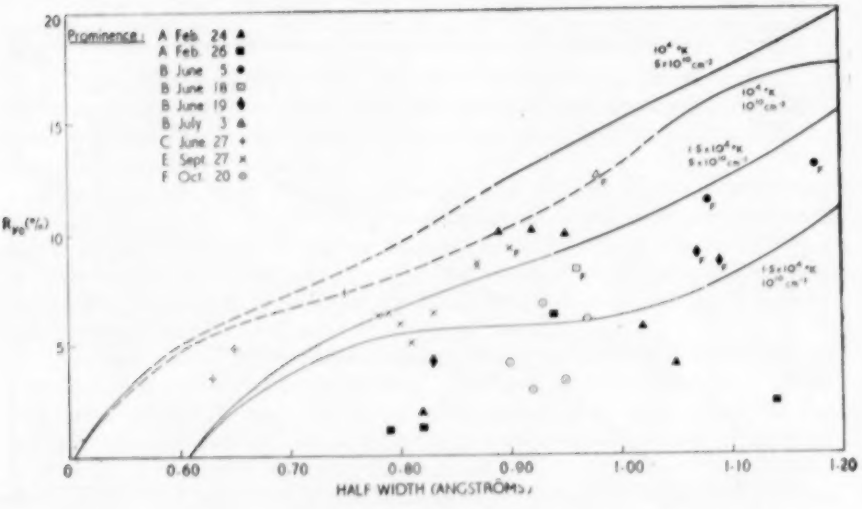
(*b*)

FIGS. 3 (*a*) *and* 3 (*b*).—*Comparison of computed central intensity–half-width curves with Conway's observations. The broken parts of the curves refer to the region of transition between low and high optical thickness in the Lyman continuum. Fig.* 3 (*a*) *refers to the coherent and* 3 (*b*) *to the partially coherent case.*

For the partially coherent case shown in Fig. 4(*b*) a higher temperature model would be required to fit some of the observations. For $T = 2{\cdot}0 \times 10^4$ deg. K, the limiting half-width is 0·72 A and the general shapes of the curves suggest that, for some prominences, temperatures may be as high as this, electron concentrations still being of the order of 10^{10} to 5×10^{10} cm^{-3}.



(*a*)



(*b*)

FIGS. 4 (*a*) *and* 4 (*b*).—*Comparison of computed central intensity–half-width curves with Ellison's observations. The heavy parts of the curves refer to computed contours with flat-topped or double-peaked profiles. The letter* F *next to an observational point indicates that this observation showed a flat or doubled profile. Fig.* 4 (*a*) *refers to the coherent,* 4 (*b*) *to the partially coherent case.*

Although the ranges of T and N_e found above seem to give agreement with observation, there is another parameter, the thickness, whose value must be considered. Each point on the curves corresponds to a certain optical thickness in Hα and so to some physical thickness of the model. Unless the thicknesses which correspond to the bulk of the observations are of the order observed for prominences, the models will be unsatisfactory. A reasonable range of thicknesses may be based on the observations of M and Mme D'Azambuja (**8**) which show that about 90 per cent of stable prominences have thicknesses between 5×10^3 and 8×10^3 km as measured on the Sun's disk. This range is rather narrow for a consideration of measurements on the limb since, owing to the filamentary structure of prominences, some observations will be made through quite small thicknesses. On the other hand, observations will not in general be made at right angles to the plane of the prominence and some allowance for this should be made. We shall adopt a thickness range of 2×10^3 to 2×10^4 km. The observations indicate that the Hα lines from a majority of prominences have half-widths in the range 0·80 to 1·00 A. If our models are to be satisfactory, those with thicknesses in the range 2×10^3 to 2×10^4 km should give half-widths about 0·80 to 1·00 A. Reference to Fig. 5(*a*) shows that in this respect the range of N_e and T seems to meet the requirements. The general trend of the curves indicates that a lower T would require a higher N_e to fit the observations; however, such a model would not fit the results of Figs. 3(*a*) and 4(*a*).

If the scattering is of the partially coherent type used here, Fig. 5(*b*) suggests that an electron concentration of 10^{10} cm^{-3} would be too low to give agreement with observation unless the temperature is about 2×10^4 deg. K or more. This would not be inconsistent with the results shown in Figs. 3(*b*) and 4(*b*).

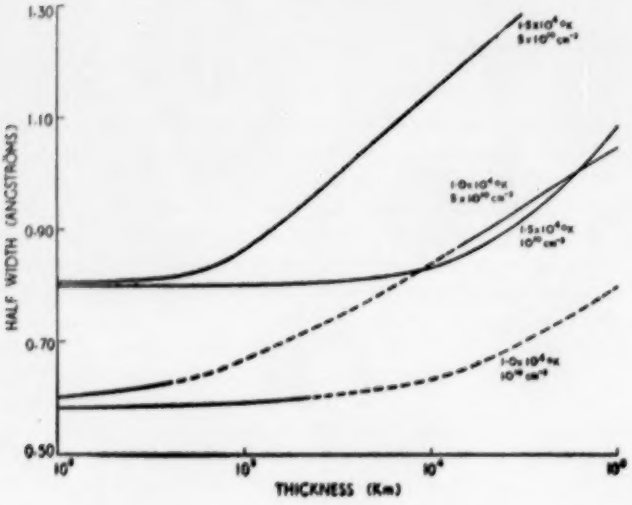
In summary, if the atomic scattering is coherent, the observational results of the Hα intensities and half-widths can only be explained by kinetic temperatures in the range about 10^4 to $1{\cdot}5 \times 10^4$ deg. K and electron concentrations about 10^{10} to 5×10^{10} cm^{-3}, the lower values of N_e being probably associated with higher T, and vice versa. For the partially coherent type of scattering adopted, the temperature range is about 10^4 to 2×10^4 deg. K, the electron concentration again being in the range about 10^{10} to 5×10^{10} cm^{-3}; the lower values of N_e are, however, consistent only with the higher values of T.

5. *Flat and double-peaked profiles.*—Ellison (**2**) describes a number of Hα prominence profiles which are flat or double-peaked. He suggests an explanation of these in terms of a prominence model consisting of a core at about 20 000 deg. K surrounded by a layer at 10 000 deg. K, pointing out that such a model could give rise to the observed profiles. While this is so, it seems more likely on physical grounds that any temperature gradient would be in the reverse direction, in which case no doubling would occur.

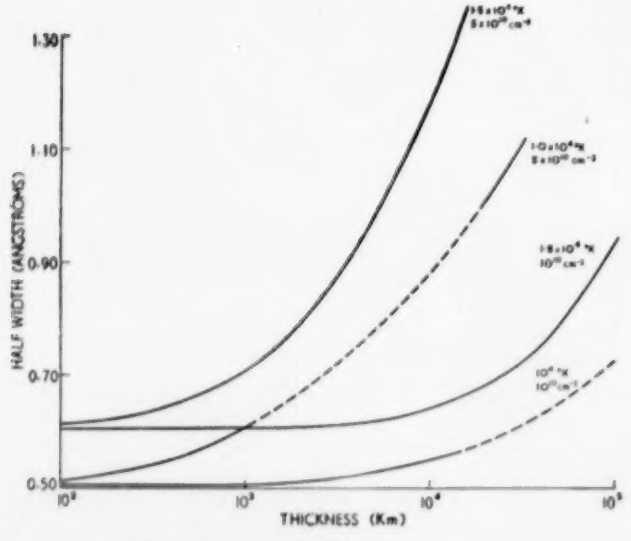
Alternative explanations arise from the present analysis. If the scattering is coherent, it is seen from (9) that, while the term representing the prominence self-emission decreases away from the line centre, the reflection term $r_\nu F(\beta, \tau_1)$ may not. Depending on the relative magnitudes of the emitted and reflected radiation, there results a normal, flat-topped or double-peaked profile as shown in Fig. 2(*a*).

For the mixed coherent and non-coherent scattering considered here, double-peaked profiles may also appear, as illustrated in Fig. 2(*b*), but for a

quite different reason. Radiation in the wings comes, on the average, from greater depths than in the centre of the line, the relative emitted intensities being proportional to the total intensities at the appropriate average depths. However, disregarding incident radiation, the total intensity increases with



(*a*)



(*b*)

FIGS. 5 (*a*) *and* 5 (*b*).—*Computed curves of half-width of the emergent radiation against thickness of the model. The broken parts of the curves, as in Figs.* 3 *and* 4, *refer to the region of transition between low and high optical depths in the Lyman continuum. Fig.* 5 (*a*) *refers to the coherent, and* 5 (*b*) *to the partially coherent case.*

depth at a rate which decreases away from the line centre. Non-coherent scattering causes some of the radiation absorbed near the line centre to be re-emitted in the wings, thus enhancing the total wing intensities and, as a result, the emitted intensity there may exceed that near the centre and a double-peaked profile result.

It follows then that for either scattering mechanism, double-peaked profiles will result if the atmosphere is sufficiently thick.

In Figs. 4(*a*) and 4(*b*) the subscript "F" indicates the flat or doubled profiles mentioned by Ellison, and the region over which the computed curves show the same behaviour is indicated by a heavier line. The agreement suggests that one of these explanations accounts for at least part of the phenomenon.

6. *Limb and disk intensities.*—Further evidence of the importance of diffuse reflection is obtained by considering the specific intensities of a prominence when observed at the limb and on the disk of the Sun. For simplicity we confine attention to an optically thick prominence; the conclusions, however, are not essentially altered by this restriction, nor, since we shall deal only with the line centre, is it dependent on the type of scattering considered, i.e. coherent or partially coherent.

The specific intensity, at the centre of Hα, of the radiation emerging at an angle $\cos^{-1}\mu$ to the normal to a thick model prominence is easily found to be given by

$$I(0,\mu)=\frac{\epsilon}{\kappa\lambda}\left[1-\frac{1-\lambda}{(1+\frac{2}{3}\beta)(1+\mu\beta)}\right]+\frac{I'}{2}\,\frac{(1-\lambda)}{(1+\frac{2}{3}\beta)(1+\mu\beta)}\,. \tag{10}$$

We shall consider two extreme cases such that (i) the reflected component is negligible, and (ii) the self-emitted component is negligible.

In case (i) it is easily seen from (10) that, when $\lambda=6\times10^{-5}$,

$$\frac{I(0,1)}{I(0,0)}\simeq\frac{5}{2(1+\beta)}=2{\cdot}2 \tag{11}$$

and so the specific intensity of a prominence observed at the limb ($\mu=1$) is over twice as great as when observed on the disk ($\mu=0$).

In case (ii) it follows simply that,

$$\frac{I(0,1)}{I(0,0)}=\frac{1}{1+\beta}=\frac{1}{1{\cdot}13} \tag{12}$$

so that the specific intensity for a limb observation is slightly less than that for a disk observation.

From observation, Ellison (**2**) found that the Hα limb and disk intensities of an individual prominence are much the same, from which it follows that diffuse reflection plays a considerable part in the Hα emission of prominences.

7. *Conclusions.*—Observational values of intensities and half-widths of the Hα emission of stable prominences have been analysed in terms of simple prominence models. It has been shown that agreement between observation and theory is obtained for models with thicknesses of the order observed and with kinetic temperatures in the range 10^4 deg. K to 2×10^4 deg. K and electron concentrations between $10^{10}\,\mathrm{cm}^{-3}$ and $5\times10^{10}\,\mathrm{cm}^{-3}$. The theoretical results have been found to be comparatively insensitive to the assumption of either fully or partially coherent scattering and the observational data are inadequate to decide between them, although on theoretical grounds the partially coherent mechanism is preferable.

It has been shown that diffuse reflection of Hα disk radiation by prominences constitutes a considerable part of their total emission and is a factor in accounting for the observed equality of limb and disk intensities of an individual prominence.

The work described in this paper was carried out as part of the programme of the Division of Physics of the Commonwealth Scientific and Industrial Research Organization, Australia.

8. *Acknowledgments.*—I should like to express my thanks to Dr R. G. Giovanelli for his constant interest and for the benefit of his valuable comments.

Division of Physics,
National Standards Laboratory,
Commonwealth Scientific and Industrial Research Organization,
Sydney :
1955 *August.*

References

(1) M. T. Conway, *M.N.*, **112,** 55, 1952.
(2) M. A. Ellison, *Pub. R. Obs. Edinburgh*, **1,** No. 5, 1952.
(3) M. T. Conway, *Proc. R. Irish Acad.*, **54 A,** 311, 1952.
(4) J. T. Jefferies and R. G. Giovanelli, *Aust. J. Phys.*, **7,** 574, 1954.
(5) S. Chandrasekhar, *Radiative Transfer*, Oxford, 1950.
(6) R. v. d. R. Woolley and D. W. N. Stibbs, *The Outer Layers of a Star*, p. 172, Oxford, 1953.
(7) J. T. Jefferies, *Aust. J. Phys.*, **6,** 22, 1953.
(8) M et Mme L. D'Azambuja, *Annales Meudon*, t. **6,** fasc. 7, 1948.

A STUDY OF THE OUTER CORONA FROM A HIGH ALTITUDE AIRCRAFT AT THE ECLIPSE OF 1954 JUNE 30

I. Observational Data

D. E. Blackwell

(Received 1955 October 21)

Summary

The eclipse of 1954 June 30 was successfully observed from an open aircraft at an altitude of 30 000 ft in excellent sky conditions. Absolute values of the coronal brightness to a distance of 13°·5 from the Sun and the coronal polarization to a distance of 5° are given, together with the sky brightness and polarization and their variation with altitude. The importance of the sky polarization is discussed. The data are related to those for the zodiacal light.

---

1. *Introduction.*—A fundamental problem in solar physics is the measurement of the brightness and polarization of the solar corona at wave-lengths in the visible spectrum, and the deduction of its electron density as a function of distance from the Sun. Many such measurements have so far been made during total eclipses of the Sun and the results up to 1945 have been surveyed by van de Hulst (**1**). Van de Hulst has used Baumbach's compilation of optical measurements (**2**), and some other material, for the derivation of a model solar corona for the equatorial and polar regions at sunspot minimum and maximum. Apart from some inconsistency in the polarization data, which we shall discuss later, this model for the minimum corona is probably reliable out to a distance of $3R_\odot$ from the solar centre. Very few reliable measurements have been made beyond $4R_\odot$, but outstanding among these are the fine measurements of Michard *et al.* (**3**) at the eclipse of 1952 February 25 at Khartoum. These authors give reasonably reliable measures of brightness to $28R_\odot$ and polarization and electron density data to $10R_\odot$. Intensity and polarization measurements between $20R_\odot$ and $60R_\odot$ have been made by Rense *et al.* from a high altitude aircraft (**4**). The significance of these measures is discussed later in this paper.

There are two clear reasons why it is difficult to make reliable observations beyond about $4R_\odot$ from the Sun. The first, which is fundamental, is that if observations are made from the ground the sky brightness at eclipse is unlikely to fall below 8×10^{-10} times the average brightness of the solar disk.* Allen† (**5**) at the 1940 eclipse recorded values of 18×10^{-10} and 12×10^{-10} in the photographic blue and red respectively, while Michard (*loc. cit.*) recorded a value of $7{\cdot}9 \times 10^{-10}$ at an effective wave-length of 6400 A in 1952. If we take the average sky brightness to be 18×10^{-10}, then van de Hulst's model shows that the coronal brightness at the equator becomes equal to this at a distance of only $4{\cdot}0R_\odot$. While it is

* Baumbach, Allen and Roach use as a unit the brightness of the centre of the solar disk, but van de Hulst and Michard use the average brightness. In this paper we use the average brightness.

† It is important to notice that these observations were made with the Sun at an altitude of 26° only.

true that it should be possible to make measurements to within 1 per cent of the sky background, it is likely that local variations of sky brightness at the average eclipse will preclude measurement to less than 10 per cent of the sky brightness. Unless conditions are exceptionally good we cannot expect to make measurements of total brightness to distances greater than $10R_{\odot}$. This difficulty can be surmounted only by making the observations at long wave-lengths or at high altitude.

The second difficulty, that of light scattering in the camera, is not fundamental and is easily overcome. But if careful attention is not paid to this point entirely spurious results may easily be obtained for distances beyond $4R_{\odot}$, even under the best atmospheric conditions.

2. *Use of aircraft.*—The observations should be made in an open aircraft, and not through an aircraft window. If they are made through an ordinary window it is very likely that the extra light scattering at the window will entirely vitiate the advantage of high altitude observations, and if a special window is fitted there is danger of frosting at high altitude. In making the observations now being described it was found most convenient to fold the aircraft door back and to mount the camera in the open doorway in such a way that it could be guided in altitude and azimuth. The aircraft used was a special Lincoln capable of flying to an altitude of 43 000 ft. Physiological difficulties made it inadvisable to attempt observations from this height, and a height of 30 000 ft was chosen to give the best compromise between sky darkening on the one hand, and aircraft stability and observer alertness on the other hand. At this height the atmospheric pressure is about 23 cm of mercury and it was hoped that the reduction in sky brightness would be by at least a factor of $\times 3$.

3. *Summary of eclipse data.*—The eclipse was observed from an altitude of 30 000 ft from the approximate position:

latitude $61^{\circ}\,23'{\cdot}9\,\mathrm{N}$

longitude $1^{\circ}\,34'{\cdot}6\,\mathrm{W}$

with the Sun practically on the meridian at an altitude of 52°. In this position the position angle of the normal to the ecliptic at the Sun was $+3^{\circ}{\cdot}7$. The position angle of the solar axis was $-3^{\circ}{\cdot}4$.

At the time of observation the aircraft was not laying a vapour trail and the sky was clear with no trace of cirrus cloud. There was dense cloud perhaps 10 000 ft below the aircraft.

A multi-lens camera, giving eight separate images on one photographic plate, was used. One photograph only was taken on an Ilford HP3 plate with an exposure time of 1 min. The exposure was begun 30 sec after the beginning of totality so that the sky background should be as dark as possible. The plate is clean and appears to be uniformly developed.

4. *Camera design.*—The camera had been designed to be as free as possible from scattered light so that full advantage could be taken of the reduced sky brightness at high altitude. It has eight simple plano-convex lenses, each of focal length 40·4 cm, which form eight separate images of the corona on a 12 inch square photographic plate. The lenses were made from Chance OY1 filter glass. The spherical aberration and coma of a simple lens is no disadvantage because the problem under examination requires no critical optical definition and because in any case it is usually not possible to guide by hand from an aircraft for a period

of one minute with an error of less than 10′. The use of a simple filter lens instead of a doublet with a separate filter also reduces spurious effects due to surface scattering and internal reflections.

Four of the lenses are used for the direct measurement of the intensity of the corona, and four for the measurement of polarization. Those for direct measurement are covered by different sized diaphragms, giving relative apertures of *f*/9, *f*/20, *f*/49, *f*/130. The three smallest lenses have total fields of $14^\circ \times 14^\circ$, and the largest has a field of $14^\circ \times 28^\circ$ with the longer dimension orientated approximately along the ecliptic.

The four lenses for the measurement of polarization form images through four pieces of Polaroid mounted about 2 cm in front of the photographic plate. The lenses are in two pairs with apertures *f*/9 and *f*/20, and each pair of Polaroids is arranged with axes parallel and perpendicular to the ecliptic.

An important part of the scattering in an eclipse camera is caused by reflection of the bright image of the inner corona from the photographic plate on to the lens, each surface of which then reflects the light back on to the plate. The effect has been reduced in this camera by blooming the surfaces of each simple lens and by obstructing the central parts of each image with velvet-faced occulting disks. When a wide angle camera is used at ground level most of the plate illumination is from the sky and the effect of a central stop is not great. Thus, if the sky brightness is $18 \times 10^{-10} \overline{B}_\odot$ and the camera field is 22°, the flux on the plate is reduced by a factor of only 1·4 by a stop of diameter $4R_\odot$. But if the sky brightness is diminished to $1{\cdot}8 \times 10^{-10} \overline{B}_\odot$, the flux on the plate is reduced by a factor of 5·0 by the same size of stop. It is particularly important that occulting disks of this kind are placed in front of the Polaroid screens because the screens reflect specularly and any uneveness in the surface produces a very uneven scattering pattern on the plate. Occulting disks used with the direct intensity images have diameters 4, $3\frac{1}{2}$, 3 and $2R_\odot$, and those used with the polarization images have diameters 4 and $3R_\odot$. To reduce further the scattering from this cause the lenses were stopped down to their correct sizes by blackened diaphragms behind as well as in front, so that a minimum area of glass is exposed to the photographic plate. The optical arrangement of one of the eight sections of the camera is shown in Fig. 1.

FIG. 1.—*Optical arrangement of one section of eclipse camera showing occulting stops A and B, Polaroid screen C and photographic plate D.*

In this camera the extra precaution is taken of eliminating images produced by secondary reflections in the lenses with little stops about 6 cm behind each lens. Scattering at the camera walls is reduced by a succession of diaphragms, and the whole coated with a dead matt black paint. Appropriate lens hoods prevent the intrusion of too much extraneous light.

The camera shutter consists of a large plate dropped in front of all eight lenses.

It will be clear that it is of the utmost importance to keep the camera lenses free of condensation. It was decided* not to attempt to heat the lenses, but to keep the air near them as dry as possible with large quantities of silica gel and to be careful not to descend from height into warm air before using the camera. These precautions proved adequate and there was no condensation.

5. *Camera tests.*—The arrangement used to test the camera for scattered light is shown in Fig. 2. Light from a 250 watt Pointolite lamp is focussed by the lens A on to the screen S, which contains a 6 mm aperture. The aperture in S

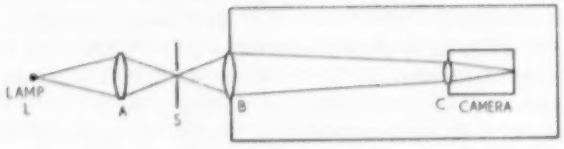


FIG. 2.—*Optical arrangement for testing eclipse camera for scattered light.*

is then focussed on to one of the camera lenses C by the lens B, about 5 cm in diameter, which is sealed into one wall of a room about 8 metres long. The camera lens forms an image of the aperture in S on the central stop, this image subtending an angle of about $\frac{3}{4}°$. With this arrangement the scattered light is easily photographed with an exposure time of $\frac{1}{2}$ min. The intensity of light falling on the lens is measured on the same plate with the same exposure time by stopping the camera lens down to pinhole size and using a neutral filter. The ratio of scattered to incident light is then evaluated by standard methods of photographic photometry.

Let E_s = total energy of scattered light per unit area of image plane,

E_0 = total energy falling on camera lens,

E_i = total energy per unit area of geometrical image,

S = coefficient of scattering.

Then S is defined by the relation

$$E_s = SE_0 = S \int_{\text{image plane}} E_i \, dA.$$

The experiment described above gives the value of S for each point in the image plane. These values may now be combined with photometric data on the solar

TABLE I

Scattering in eclipse camera

R	S	P
$4{\cdot}9\ R_\odot$	$18{\cdot}5 \times 10^{-6}$	200
6·7	8·1	290
8·2	4·0	360
11·5	1·4	780
14·8	0·77	1200

* On the advice of Mr N. P. Court of the Royal Aircraft Establishment, Farnborough.

corona to give P, the ratio of the intensity of the corona and sky at each point to the intensity of the scattered light which, in these tests, is chiefly caused by the imperfect surface polish of the lenses. Some data are given in the following table for several distances from the Sun. The sky brightness has been taken to be, for the sake of example, $10^{-10}B_{\odot}$. The table shows that the lens surfaces are not of the highest quality, but as the scattered light is never more than 0·5 per cent of the coronal light and less than 0·1 per cent at great elongations, they are amply good enough.

6. *Plate calibration.*—Immediately after the eclipse, at the same altitude, another photographic plate from the same batch as was used for the eclipse photograph was impressed with calibration marks. This was done in a large box near the eclipse camera at one end of which was the standard lamp and Chance filter, and at the other end, at a distance of 5 ft, the photographic plate mounted behind five gelatin wedges. The same exposure time as for the eclipse was used. This plate was afterwards fastened with the eclipse plate into a large frame and the two developed together, with vigorous brushing, in Ilford ID-2 developer. They were then fixed after immersion in a stop bath.

The five wedges were later calibrated in the same apparatus against a platinum step wedge which had been standardized at the National Physical Laboratory. A self calibration could also be obtained from the four coronal images themselves, for the intensities of these are simply related, and it is satisfactory that these calibrations agree over most of the characteristic curve to within 1 per cent.

The standard lamp was a 4 volt 4 watt Royal Ediswan Vacuum hairpin filament lamp, which was standardized at the National Physical Laboratory. Its colour temperature was 1490 deg. K. The data can therefore be used, in a fairly straightforward manner, to determine the absolute intensity of radiation from the solar corona. There is, however, a slight difficulty introduced by our complete ignorance of the colour temperature of the corona and sky. This does not matter if the sensitive wave-length range is small, but in these observations it is about 700 A, and the final result must, to a certain extent, depend upon the assumption made about the colour temperature of the corona.

Let S_λ be proportional to the product of plate sensitivity and filter transmission as a function of wave-length. Then, if E_λ(lamp) is the energy curve of the standard lamp, the energy falling on the photographic plate in the calibration box is proportional to

$$\int_{\text{all }\lambda} S_\lambda E_\lambda(\text{lamp})\, d\lambda.$$

If at one point in the coronal photograph the density is the same as the density of the calibration plate, then

$$\int_{\text{all }\lambda} S_\lambda E_\lambda(\text{corona})\, d\lambda = \int_{\text{all }\lambda} S_\lambda E_\lambda(\text{lamp})\, d\lambda.$$

These integrals are the areas under the curves $S_\lambda \times E_\lambda$. The rate of radiation from the corona has now been found in the following manner. First, a likely colour temperature for the corona is assumed and then the curves $S_\lambda \times E_\lambda$ are plotted on an arbitrary scale for the corona and the standard lamp. The areas beneath these curves are measured and the ordinates scaled until the areas are the same. The diagram, Fig. 3, shows the two curves scaled to equal areas. It will be seen that the 6000 deg. K curve emphasizes the shorter wave-length

parts in comparison with the larger wave-lengths. At the intersection of the curves, point P, S_λ is the same for both sources, and hence E_λ is the same (the total energy is the same because the areas are the same). We define the wave-length of the intersection of the two curves as the "effective wave-length", and the rate of radiation from the corona at this wave-length is equal to the rate of radiation from the standard lamp at the same wave-length. The effective wave-length is very close to 6300 A.

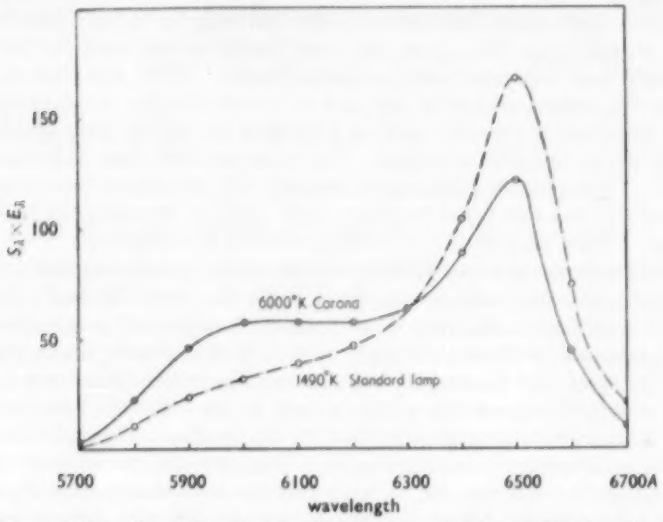


FIG. 3.—*Spectral sensitivity of camera for sources of colour temperature 6000 deg. K and 1490 deg. K.*

This analysis is of little use if the effective wave-length varies greatly with the assumption about the colour temperature of the corona. In fact, a repeat of this procedure for colour temperatures of 4000 deg. and 8000 deg. shows that the effective wave-length does not change by more than $\pm$ 10 A and the deduced absolute energy at this wave-length is uncertain by no more than 1·5 per cent. This is well within the range of experimental error.

The variation of sensitivity with wave-length of the photographic plate was measured afterwards on the ground with a prism spectrograph and a source of known colour temperature. It may be that this curve differs sensibly from what would have been obtained from a similar experiment in the air at a much lower temperature, but the determination of the absolute energy from the corona is very insensitive to the precise form of S_λ. What matters is that the calibration photograph and the eclipse photograph were both made under identical conditions.

The observations thus give absolute values for the rate of radiation of energy from various points of the projected corona. These values have also been expressed in terms of the mean brightness of the Sun at the effective wave-length by interpolating the data given by Minnaert (**6**).

At the time of observation the sky was perfectly clear with no trace of cirrus cloud. No correction has therefore been made for atmospheric absorption. This is justified by reference to the Potsdam extinction tables (**7**). These tables

show that at a height of 8200 ft and an altitude of 52° the extinction at a wave-length 5300 A is 3·8 per cent. Hence, the extinction at a height of 30 000 ft and a wave-length of 6300 A is of the order of

$$3{\cdot}8 \times \left(\frac{0{\cdot}53}{0{\cdot}63}\right)^4 \times \frac{\text{atmos. pressure at } 30\,000 \text{ ft}}{\text{atmos. pressure at } 8200 \text{ ft}}$$

$$\simeq 0{\cdot}5 \text{ per cent.}$$

This can be completely neglected.

7. *Effect of aircraft movement.*—An investigation has been made of the effect of image movement, due to imperfect guiding, on the photometric observations.

During the exposure the aircraft remained remarkably steady, not deviating from its undisturbed position by more than ±8′. Hand guiding of the camera about horizontal and vertical axes reduced this error by a factor of 2 to ±4′.

The effect of imperfect guiding is to increase the measured intensity of the corona everywhere except very close to the occulting disk. The intensity is increased because, as the curve is convex to the axis, with an image movement XY the gain in intensity AB is greater than the loss AC (see Fig. 4). The total intensity, integrated over all distances from the Sun, remains the same; that is, the areas α and β are the same.

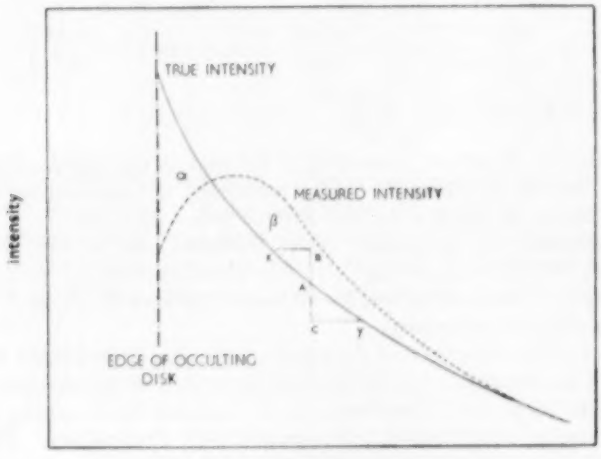


FIG. 4.—*Effect of lens aberration and errors of guiding on the photometric accuracy. The effect is greatly exaggerated in this diagram.*

The calculation of the correction depends upon the nature of the image movement. An unsymmetrical movement about the mean position would give rise to large systematic errors because of the uncertain position of the Sun. To avoid these extreme errors the intensities of the four quadrants have been averaged.

If we now suppose the image movement to be symmetrical about the position of the Sun, the most serious error will occur if the image divides its life-time equally between the points X and Y

Suppose that, at a particular point, the true intensity of the corona varies with distance according to the relation

$$I_e = kr^{-n}.$$

Then, if the image movement at distance R is $\pm\Delta r$, the measured intensity will be given by

$$I_{\text{meas}} = \tfrac{1}{2}kR^{-n}\left[\left(1+\frac{\Delta r}{R}\right)^{-n} + \left(1-\frac{\Delta r}{R}\right)^{-n}\right]$$
$$= I_e\left[1+\frac{n(n+1)}{2}\left(\frac{\Delta r}{R}\right)^2\right].$$

Hence the correction is equal to

$$I_e \times \frac{n(n+1)}{2}\left(\frac{\Delta r}{R}\right)^2.$$

The example in which the image oscillates uniformly between the two extremes is more favourable. Then, if the oscillation is between $R-\Delta r$ and $R+\Delta r$, the measured intensity is

$$I_{\text{meas}} = \int_{R-\Delta r}^{R+\Delta r} I_e\,dr/2\Delta r$$
$$= \frac{kR^{1-n}}{2\Delta r(1-n)}\left[\left(1+\frac{\Delta r}{R}\right)^{1-n} - \left(1-\frac{\Delta r}{R}\right)^{1-n}\right]_{R-\Delta r}^{R+\Delta r}$$
$$= \frac{kR^{1-n}}{2\Delta r(1-n)}\left[2(1-n)\left(\frac{\Delta r}{R}\right) - \frac{1}{3}n(n-1)(n+1)\left(\frac{\Delta r}{R}\right)^3\right]$$
$$= I_e\left[1+\frac{1}{6}n(n+1)\left(\frac{\Delta r}{R}\right)^2\right].$$

The correction is therefore $\frac{1}{6}I_e n(n+1)(\Delta r/R)^2$ and is one third of the former correction. As the nature of the image movement is unknown we shall form a good estimate of the error if we take it to be $\frac{1}{3}n(n+1)(\Delta r/R)^2$.

The correction due to guiding errors increases rapidly with diminishing distance from the Sun because both n and $\Delta r/R$ are increasing. At $5R_\odot$ it is of the order of 0·2 per cent and is less for all greater distances from the Sun. This correction is therefore neglected.

The effect of the aberration of the camera lenses has been treated in a similar fashion, with the result that for all distances from the Sun greater than $5R_\odot$ the error caused by its neglect is less than 0·01 per cent and can be wholly neglected.

8. *Reduction of direct intensity photographs.*—(i) Preliminary. The photographs have been measured with the Observatories' microphotometer. This was originally a Cambridge Instrument Co. microphotometer and has been completely rebuilt in the Observatories' workshops to the design of Professor R. O. Redman and Mr G. G. Yates. The output from the photoelectric cell, which is proportional to the transmission of the plate, is amplified by a d.c. amplifier and fed into a pen recorder. The images were scanned with a wide slit to reduce the effect of plate grain. The slit size on the photographic plate was 0·36 mm × 0·10 mm which is equivalent to $0{\cdot}2R_\odot \times 0{\cdot}05R_\odot$.

(ii) The shadows of the bars supporting the occulting disks have been used as reference axes for the photographs. During the observations these bars were horizontal and vertical to the limit of accuracy with which the aircraft was held level; the tilt was measured immediately before the eclipse with a spirit level.

As the bars almost exactly bisect the angle between the solar equator and the ecliptic, and these directions are for the most part hidden behind the bars, both have, with negligible error, been taken to coincide with the bars (see Fig. 7).

(iii) Sky brightness. This has been found by making microphotometer scans of the photographs perpendicular to the ecliptic. Straight lines have been drawn touching the curves at the upper and lower ends, and it has been supposed that these lines give the intensity of the sky and its variation with altitude (see Fig. 5).

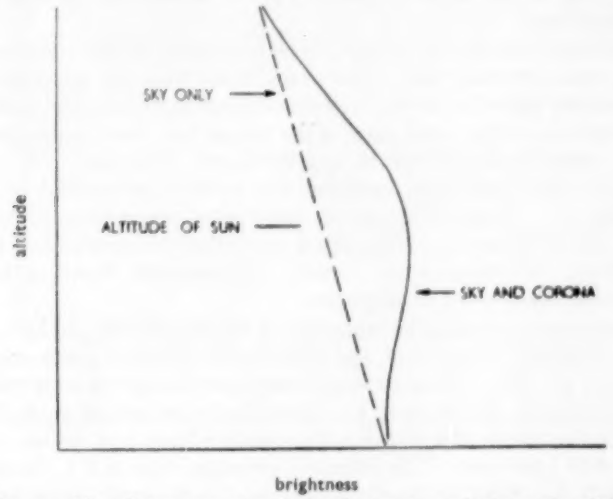


FIG. 5.—*Microphotometer tracing obtained by scan perpendicular to the ecliptic.*

Over the field of view of the camera the sky brightness is independent of azimuth, but markedly dependent on altitude. The variation with altitude is given in the accompanying table, where the unit of brightness is the average brightness of the solar disk at a wave-length in the neighbourhood of 6300 A (see p. 634).

TABLE II

Sky Brightness

Altitude	Sky brightness
58°	$2{\cdot}48 \times 10^{-10}\,\overline{B}_{\odot}$
52°	2·64
46°	2·80

It is interesting to compare these values for the sky brightness with those found at Khartoum by Michard *et alia* for the 1942 eclipse ($7{\cdot}90 \times 10^{-10}\overline{B}_{\odot}$). The effective wave-length used by Michard was 6400 A and if we make a correction for the small difference in wave-length we find that the ratio of the ground brightness in 1952 to the brightness at 30 000 in 1954 is

$$\frac{7{\cdot}90}{2{\cdot}64} \times \left(\frac{64}{63}\right)^4 = 3{\cdot}2.$$

The atmospheric pressure at 30 000 ft is 23 cm mercury and the ratio of this pressure to ground pressure is approximately 3·3. The observed ratio of sky brightness is slightly smaller than this, whereas, because of the purity of the upper air, we would have expected it to be rather greater. The discrepancy probably arises from the fact that in 1952 the duration was 189 sec and the solar altitude 62° against quantities of 155 sec and 52° for the 1954 eclipse.

From these data we deduce that the brightness of the eclipse sky varies with the height of the observer approximately in proportion to the atmospheric pressure, and further that the sky at the eclipse of 1952 at Khartoum was nearly as dark as the practical limit.

(iv) The brightness of the corona. The brightness of the inner part of the corona has been obtained from radial scans made with the microphotometer, a correction being made for the sky brightness according to the position in the sky.

The brightness of the outer parts of the corona has been found from microphotometer traces made in vertical and horizontal directions. All deviations from a straight line have been supposed due to the superposition of the solar corona on the sky. These deviations are easily measured and converted into line intensities. In the final presentation of the results the average of all four quadrants has been given. Measures in the vertical and horizontal directions have been made out to the edges of the photographs.

The relative intensities in the inner part of the corona ($5R_\odot$ to $25R_\odot$) should be accurate to within ± 5 per cent, and those in the outermost parts of the corona to within ± 10 per cent. No greater accuracy than this can be expected because at $54R_\odot$, for example, the intensity is no more than 2 per cent of the sky intensity. The error in the comparison between the standard lamp and the corona should be no more than 5 per cent. The estimated accuracy of the N.P.L. determination of the absolute flux from the standard lamp is ± 10 per cent, giving an accuracy of ± 12 per cent for the absolute brightness of the corona.

The bars supporting the occulting disks have been placed vertical and horizontal and it has been necessary to make an extrapolation across them to obtain the brightness of the corona in the equatorial regions (see Fig. 7). The bars have a width on the plate of 0°·5.

Data for the measures along the ecliptic are given in Fig. 6 and a model corona* obtained by smoothing the observations of Fig. 6, is given in Table III. In Fig. 6 there are also plotted the data of Michard from the 1952 eclipse and the coronal model of van de Hulst. It will be seen that the curve for 1954 lies between these two, and is very close to van de Hulst's model. The absolute intensity is greater than that of the model by a factor of × 1·29 at $10R_\odot$, and the gradient between $6R_\odot$ and $10R_\odot$ coincides with that of the maximum model to within 1 per cent.†

There is a logarithmic plot of the results in Fig. 10, which shows that they can be represented approximately by a relation of the form,

$$I = kr^{-n}.$$

For values of r between $6R_\odot$ and $55R_\odot$, the best fit with the data is obtained for

$$n = 2{\cdot}20 \text{ (equator)}.$$

---

* Note that the solar equator is at P.A. = 90°.

† There is only a small difference between the maximum and minimum distributions at distances greater than $6R_\odot$ from the Sun.

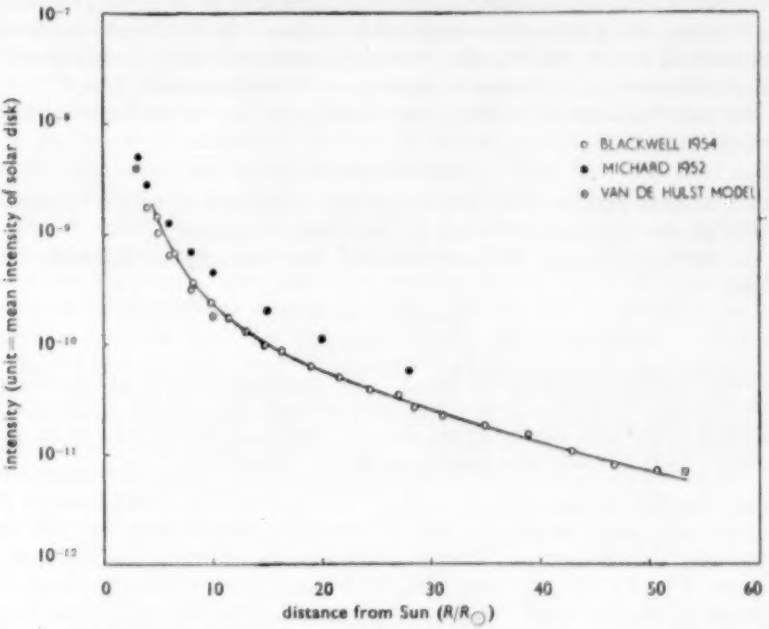


FIG. 6.—*Brightness variation of solar corona along ecliptic (which is practically coincident with solar equator).*

TABLE III

Brightness of corona

Elongation	Position Angle			Log brightness of corona (unit $10^{-13}\,\bar{B}_\odot$)				
$R/R_\odot$	90°	75°	70°	60°	45°	30°	20°	0°
6	2·93	2·87	2·84	2·78	2·74	2·72	2·71	2·70
7	2·76	2·68	2·65	2·60	2·54	2·51	2·50	2·49
8	2·61	2·54	2·51	2·45	2·38	2·34	2·33	2·32
9	2·47	2·41	2·38	2·32	2·26	2·21	2·19	2·18
10	2·37	2·31	2·28	2·21	2·13	2·09	2·07	2·06
11	2·28	2·20	2·17	2·10	2·03	1·98	1·96	1·95
12	2·20	2·10	2·06	1·99	1·92	1·87	1·85	1·85
14	2·06	1·94	1·90	1·83	1·76	1·74	1·73	1·72
16	1·94	1·81	1·77	1·69	1·64	1·63	1·63	1·63
18	1·84	1·68	1·64	1·58	1·56	1·52	1·52	1·52
20	1·76	1·58	1·53	1·46	1·43	1·42	1·42	1·42
25	1·57	1·33	1·29	1·24	1·21			
30	1·40	1·06	1·03	0·99	0·97			
35	1·25	0·90	0·83	0·76	0·73			
40	1·10							
45	0·97							
50	0·84							
55	0·72							

This relation then reproduces the observed results with an error not greater than 11 per cent. But it is emphasized that this relation is no more than a convenient mathematical device, and that the observed deviations from it clearly shown in Fig. 10 represent real deviations and not errors of measurement.

Between the limits of $r = 6R_\odot$ and $r = 20R_\odot$ the best fit for the polar data is obtained for

$$n = 2{\cdot}47 \text{ (pole)}.$$

This is a mean gradient only which reproduces the data to within 12 per cent but for greater accuracy individual points should be plotted.

At the 1952 eclipse Michard obtained the following considerably lower values,

equator $n = 2{\cdot}0$

pole $n = 2{\cdot}23$.

Coronal isophotes are shown in Fig. 7. The isophotes do not have an elliptical form but tend to have a spiked structure on each side, which is obviously a remanent of the well marked spikes seen on published photographs of the inner corona. Each isophote is an average of all four quadrants.

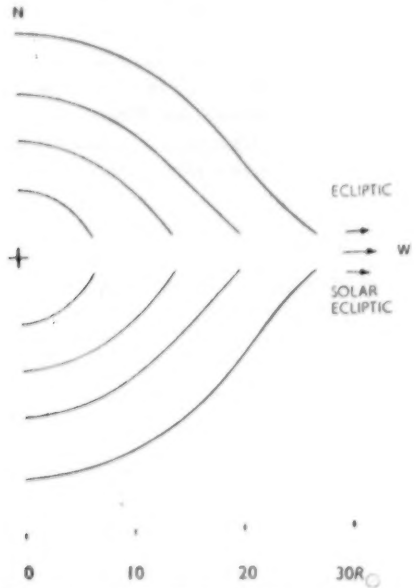


FIG. 7.—*Isophotes of the outer corona. Each isophote is an average for the four quadrants.*

(v) The polarization of the corona. The degree of polarization is obtained by comparing the intensities of pairs of images photographed through Polaroid screens with their axes at right angles to each other. If the direction of polarization is assumed*, the degree of polarization can be calculated. The direction of polarization of the coronal light is always assumed to be perpendicular to the radius vector—and it would be difficult to give theoretical justification for any other direction. But before the coronal polarization can be measured it is of the

* The direction of polarization is defined to be the direction of the electric vector.

utmost importance to measure the degree of polarization and the direction of polarization of the sky background. The measurement of two images will not give both of these quantities independently, but from the nature of the measurements it is very likely that the electric vector is predominantly vertical for small distances from the Sun. This assumption is justified by the fact that, if it is made, the degree of polarization is found to be independent of azimuth distance from the Sun, although dependent upon altitude. Now it is clear that under the good sky conditions of the present eclipse, only a small proportion of the sky background is due to scattering of the inner corona*, and most of it is due to secondary scattering of light from outside the eclipse cone. For reasons of symmetry, therefore, the electric vector could be expected to be predominantly vertical near to the Sun. Having settled this point we can proceed to calculate from the measurements the sky polarization. It will be shown later that it is possible to measure the polarization of the sky background without assuming that the polarization of the F corona is zero. To give a preliminary orientation to the problem we quote at this stage the measured values of the sky background polarization. These are the averages of all measurements on the two pairs of polarization photographs. Measurements on the two pairs are completely independent but there is good agreement between them. Thus one pair gives for the polarization at 52° altitude the value 15·2 per cent and the other pair gives 15·4 per cent with a mean of 15·3 per cent. These results are in good agreement with those of Michard (*loc. cit.*) who found a value of 12 per cent. The difference is probably due to the lower altitude of the Sun at the 1954 eclipse. At the eclipse of 1945 July 9, with the Sun at an altitude of 39°, Öhman (**8**) observed a polarization of about 30 per cent. It is reasonable to suppose that the polarization is independent of the height of the observer.

TABLE IV

Polarization of sky background

Altitude	Polarization
57°	13·3 per cent
52°	15·3
47°	17·3

As the polarization of the corona in the region of $20R_\odot$ from the Sun is of the order of 3 per cent only, and as, in addition, the coronal light is diluted with sky light by a factor of at least × 5, it is evident that we must proceed very cautiously. In Fig. 8 consider a point P in the neighbourhood of the Sun, the direction SP making an angle θ with the horizontal. The axes of the two Polaroids of a pair are orientated horizontally and vertically along the directions (2) and (1). The total sky intensities transmitted through the two Polaroids are I_{s_2} and I_{s_1} respectively. Also K_t and K_r are the tangential and radial components of the coronal light intensity. Then, if I_2 and I_1 are the total intensities transmitted by the two Polaroids, we have,

$$I_1 = I_{s_1} + K_t \cos^2\theta + K_r \sin^2\theta$$
$$I_2 = I_{s_2} + K_t \sin^2\theta + K_r \cos^2\theta.$$

* The contribution to the sky background from this cause will not be greater than $10^{-13}\bar{B}_\odot$.

The observed polarization of the coronal and sky combined

$$= P_{\text{meas}} = \frac{I_1 - I_2}{I_1 + I_2}$$

$$= \frac{(I_{s_1} - I_{s_2}) + \cos 2\theta (K_t - K_r)}{(I_{s_1} + I_{s_2}) + (K_t - K_r)}.$$

The total sky intensity $= I_s = I_{s_1} + I_{s_2}$ and the total coronal intensity $= I_K = K_t + K_r$. Hence

$$P_{\text{meas}} = \frac{I_{s_1} - I_{s_2}}{I_s + I_K} + \frac{K_t - K_r}{I_s + (K_t + K_r)} \cos 2\theta.$$

Put $I_K / I_s = K.$

Then $$P_{\text{meas}} = \frac{I_{s_1} - I_{s_2}}{I_s(1+K)} + \frac{K_t - K_r}{(K_t + K_r)[1 + (1/K)]} \cos 2\theta$$

or $$P_{\text{meas}} = \frac{P_{\text{sky}}}{1+K} + P_{\text{corona}} \times \frac{\cos 2\theta}{1 + (1/K)}$$

where

P_{sky} = sky polarization

and

P_{corona} = coronal polarization.

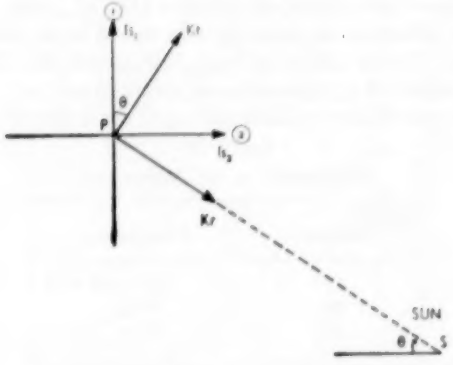


FIG. 8.—*Combination of sky and coronal polarizations.*

This result shows that the contribution of the sky to the observed polarization is $P_{\text{sky}}/(1+K)$ and the contribution of the corona is $P_{\text{corona}} \times \cos 2\theta / [1 + (1/K)]$. The coronal polarization is calculated from the total polarization measured on the plate using the relation

$$P_{\text{corona}} = \left(P_{\text{meas}} - \frac{P_{\text{sky}}}{1+K} \right) \frac{1 + (1/K)}{\cos 2\theta}.$$

We notice now that the accuracy of the final result depends upon the accuracy with which the quantity in brackets can be measured. Thus at $20 R_\odot$ from the Sun, where the polarization of the corona on the equator is about 2·8 per cent, if the sky background is $7{\cdot}9 \times 10^{-10} B_\odot$ the quantity in brackets is only 0·25 per cent. This is beyond the reach of photographic photometry. Table V gives values of $P_{\text{meas}} - P_{\text{sky}}/(1+K)$ for various values of sky brightness and distance from the

Sun, using coronal polarization data derived from the present series of measurements and tabulated in Table VII. This table gives the actual values of the percentage polarizations that have to be measured.

TABLE V

Limitation of accuracy of polarization measurements imposed by sky background

$\frac{R}{R_\odot}$	Assumed coronal polarization	Values of $\left(P_{meas} - \frac{P_{sky}}{1+K}\right)$ for various assumed values of sky brightness		
		$18 \times 10^{-10}\,\overline{B}_\odot$ (average good sky)	$7{\cdot}9 \times 10^{-10}\,\overline{B}_\odot$ (Michard 1952)	$2{\cdot}64 \times 10^{-10}\,\overline{B}_\odot$ (D.E.B. 1954)
5	9·4 per cent	4·1 per cent	6·0 per cent	7·9 per cent
10	4·3	0·49	0·98	2·02
15	3·1	0·16	0·35	0·85
20	2·8	0·11	0·25	0·63

It will also be clear that to obtain an accurate value of the coronal polarization the sky polarization must be accurately known. It follows from the above formula that an error ΔP_{sky} in the sky polarization will lead to an error ΔP_c in the coronal polarization where

$$\Delta P_c = A \times \Delta P_{sky}$$

and

$$A = \frac{1+(1/K)}{1+K} \cdot \frac{1}{\cos 2\theta}.$$

Values of A are tabulated in Table VI for various values of sky brightness. This table shows that if we wish to have an absolute accuracy of 0·35 per cent in the measurement of coronal polarization at $20R_\odot$, then the sky polarization must be known to within 0·015 per cent, 0·35 per cent and 0·13 per cent for the three values of sky brightness.

TABLE VI

Limitation of accuracy of polarization measurement imposed by sky background

$R/R_\odot$	Values of A corresponding to sky brightness		
	$18 \times 10^{-10}\,\overline{B}_\odot$	$7{\cdot}9 \times 10^{-10}\,\overline{B}_\odot$	$2{\cdot}64 \times 10^{-10}\,\overline{B}_\odot$
5	1·28	0·56	0·19
10	7·7	3·4	1·7
15	18	7·9	2·6
20	24	10	3·5

The two tables show how important it is, in the measurement of polarization, to have the sky background as dark as possible. They also show that under the best atmospheric conditions at ground level it is not feasible to measure the coronal polarization beyond about $10R_\odot$ from the Sun, or beyond about $20R_\odot$ at a height of 30 000 ft.

In measuring the polarization photographs particular care was taken to avoid confusion between, on the one hand, a slight differential variation with position of plate sensitivity or Polaroid transparency and, on the other, a variation of polarization with position. Instrumental errors of this kind were investigated by making microphotometer scans along lines parallel to the ecliptic and further than 3° from it, where (as the contribution from the corona is less than 0·1 per cent) all variations of measured polarization are instrumental. A second method used for investigating these errors made use of the neutral zones of the Polaroid. Along the lines where $\theta = 45°$ and 135° the solar corona makes no contribution to the observed polarization but acts only as a diluent of the sky polarization. Hence, as the sky polarization is known (from measures far from the Sun) and the ratio of the coronal intensity to sky background intensity is also known the instrumental error can be evaluated along these lines.

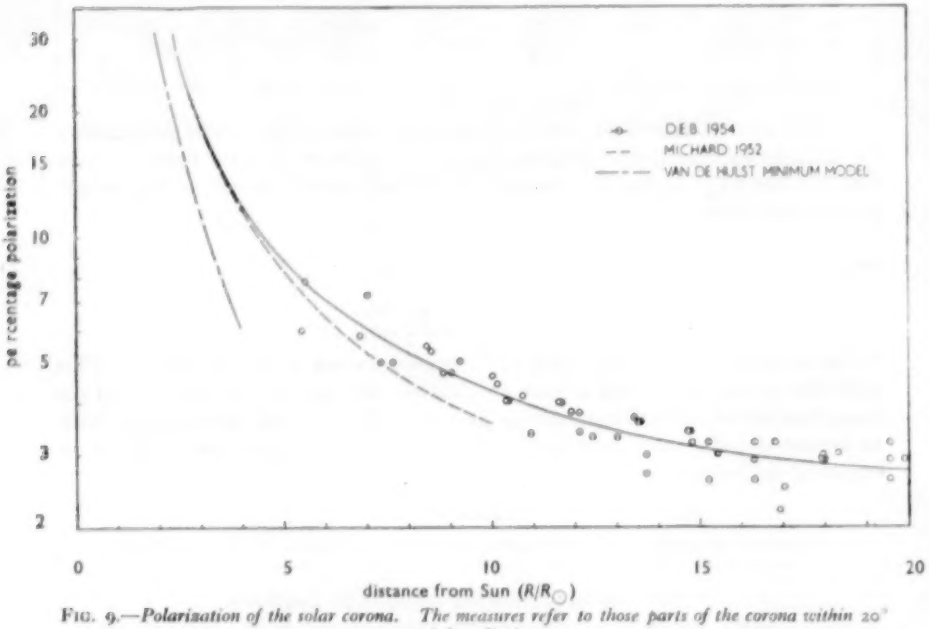


FIG. 9.—*Polarization of the solar corona. The measures refer to those parts of the corona within 20° of the ecliptic.*

These observations have given the polarization of the corona in the region of the ecliptic to a distance of $20R_\odot$, this being the field of the polarization photographs. The absolute accuracy of the measurement at $20R_\odot$ is about 0·2 per cent. Individual points, each the mean of eight measures derived from four quadrants in two independent pairs of photographs, are given in Fig. 9 and a smoothed model is tabulated in Table VII. The agreement with Michard's data is reasonably good, the discrepancy at $10R_\odot$ being only 0·7 per cent of polarization. The polarizations of the model of van de Hulst, however, seem to be much too low. Part of the spread of observations shown in Fig. 9 is genuine because there are included in this diagram all measures within the position angle range of 20° from the ecliptic. In particular, this accounts for the large spread close to the Sun.

Some measures have been made near the poles, but the present observations are not satisfactory in this region. One reason for this is that at sunspot minimum the polarization is sufficiently large to be measurable only near the Sun, and the photographs, intended for the outer regions, are rather over-exposed here. But

TABLE VII

Polarization of the corona (equatorial region)

$R/R_\odot$	$p(\%)$	$R/R_\odot$	$p(\%)$
3·0	19·0	9·5	4·5$_6$
3·5	14·5	10·0	4·3$_5$
4·0	12·0	11·0	4·0$_0$
4·5	10·2$_8$	12·0	3·7$_0$
5·0	9·0$_0$	13·0	3·4$_4$
5·5	7·9$_4$	14·0	3·2$_4$
6·0	7·1$_2$	15·0	3·0$_8$
6·5	6·5$_0$	16·0	2·9$_7$
7·0	6·0$_4$	17·0	2·8$_8$
7·5	5·6$_6$	18·0	2·8$_2$
8·0	5·3$_2$	19·0	2·7$_9$
8·5	5·0$_4$	20·0	2·7$_5$
9·0	4·8$_0$		

the problem of the polarization at the poles does not need high altitude observations for its solution because, so close to the Sun, there is only a small sky correction even from the ground. The few observations made are given in Table VIII. We wish to emphasize that in deducing these results it has not been necessary to make any assumption about the state of polarization of the F-corona.

TABLE VIII

Polarization of the corona (polar region)

$R/R_\odot$	Polarization
3·3	3·5 per cent
4·9	1·0
6·6	0

9. *Effect of sky background on coronal polarization measures near the Sun.*— Several methods have been used for separating the F and K coronas. In one method the intensities of the Fraunhofer lines in the coronal spectrum, supposed produced by dust diffraction alone, are compared with their intensities in the solar spectrum. In another method it is assumed that the F-corona is unpolarized, so that a separation is achieved by measurement of the total coronal polarization. A third method (**9**) makes assumptions about the gradient and intensity of the K-component at sunspot maximum and minimum. There have been serious discrepancies between the results obtained by these methods, the computed polarization of a coronal model deduced by the third method being always less than the observed polarization. The cause of this discrepancy probably lies in neglect of the background polarization, and it is very important to attempt a quantitative explanation of it because measures of electron density in the outer corona depend almost entirely on the measurement of polarization.

To illustrate this we take Allen's measures of polarization at the 1940 eclipse (**10**). His measurement at $3{\cdot}7R_\odot$ from the solar centre is (neglecting the sky background) of the order of 43 per cent in the blue region, whereas from van de Hulst's model we would expect a polarization of about 12 per cent.

According to Allen's measures the intensity of the sky background is $18\times10^{-10}\bar{B}_\odot$. Hence, at $2{\cdot}7R_\odot$, K has the value 1·2.

The altitude of the eclipsed Sun was 26°, and we will suppose that the polarization of the sky background was 40 per cent (this is compatible with the present results, and those of Michard and of Öhman).

Then we would measure on the photographic plate a value

$$P_{\text{meas}}=\frac{40}{2{\cdot}2}+\frac{12}{1{\cdot}83}=24{\cdot}8 \text{ per cent}$$

assuming that the coronal polarization is 12 per cent. If this measurement is now interpreted with $P_{\text{sky}}=0$ (i.e. a dilution correction only is made), then P_c would be found equal to $(1+(1/K))P_{\text{meas}}=1{\cdot}83\times24{\cdot}8=45$ per cent. This value is very close to that of 43 per cent found by Allen. We surmise therefore that the discrepancy arises from neglect of sky polarization, and that when this is taken into account the coronal polarization may be used for the deduction of reliable electron densities.

10. *Comparison with the observations of Rense et al.*—These observations were made from an aircraft at a height of 32 000 ft. It is not easy to compare the present results concerning the brightness of the corona with those of Rense *et al.*, because although Rense does give absolute values for the total sky brightness he does not give the range of spectral sensitivity of the camera. Also, the method used by Rense for correcting the observations for the presence of the sky background is a most uncertain one.

The value given by Rense for the polarization of the corona—20 per cent (corresponding to a depolarization factor ρ of 0·80) at an elongation of 5°·5—is an order of magnitude different from the value found in the present investigation 2·7 per cent at 5°·0. Two comments about this comparison should be made here. First, the discrepancy between the observations probably arises from the method of allowing for the polarization of the background sky. Second, if the polarization of the corona is really as low as is indicated in Table VII, it is not practicable to measure it against the sky by photographic photometry at this height to a distance greater than 5° from the Sun.

11. *Comparison between coronal data and zodiacal light data.*—The question of whether the zodiacal light can be considered to be part of the solar corona has been frequently discussed, especially in recent years. To demonstrate the connection between the two it is of interest to follow van de Hulst and plot on the same diagram the brightness of the corona and zodiacal light as a function of distance from the Sun. This is done in Fig. 10 using the zodiacal light data of Roach *et al.* (**11**). It will be seen that the two curves are continuous, although it should be noted also that the zodiacal light observations have been made at an effective wavelength of 5300 A.* There still remains between the furthest observable part of the corona and the nearest observable part of the zodiacal light a distance gap of $\times2{\cdot}2$ and an intensity gap of $\times6{\cdot}3$.

* The zodical light intensity is different for the E and W directions and is dependent on the season. The values used are the mean values given by Roach *et al.*

The mean exponent n in the relation $I = kr^{-n}$, given previously, agrees well with the value derived by Roach for the zodiacal light. The two values are

solar corona in the region of the ecliptic, which at this time coincided with the equatorial region of the Sun

$$n = 2{\cdot}20$$

zodiacal light

$$n = 2{\cdot}22.$$

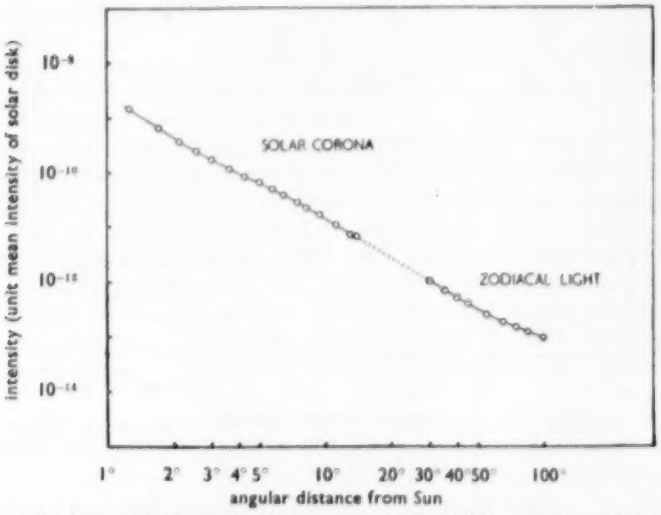


FIG. 10.—*Variation of brightness of solar corona and zodiacal light with distance from the Sun. The zodiacal light data are taken from the work of Roach et al.*

The value given by Michard, $n = 2{\cdot}0$, agrees less well, but this may be due to a real fluctuation in the intensity of the outer corona. According to current theories the F corona is caused by the diffraction of sunlight by those particles which lie between Earth and Sun. A local concentration of particles at one position within the Earth's orbit could cause a change in the brightness and gradient of the outer corona.

Mean polarizations are plotted in Fig. 11, the coronal data being averaged from Fig. 7 and the zodiacal light data being that of Behr and Siedentopf (**12**), referring to a wave-length of 5420 A. In interpreting this diagram it should be borne in mind that the distance scale is logarithmic. The two curves must be continuous and indicate a minimum polarization at about 6° from the Sun.

The Ludendorff flattening index is plotted in Fig. 12. If for any isophote d_1 is the mean of the maximum diameter and the diameter which make an angle of $22\frac{1}{2}°$ on each side of it, and d_2 is the mean of the minimum diameter and the diameters which make an angle of $22\frac{1}{2}°$ on each side of it, then the flattening index is defined by the relation

$$\epsilon = \frac{d_1}{d} - 1.$$

The points for the inner corona are taken from the work of van de Hulst, using his model corona (Kuiper, p. 287). The points for the zodiacal light have been

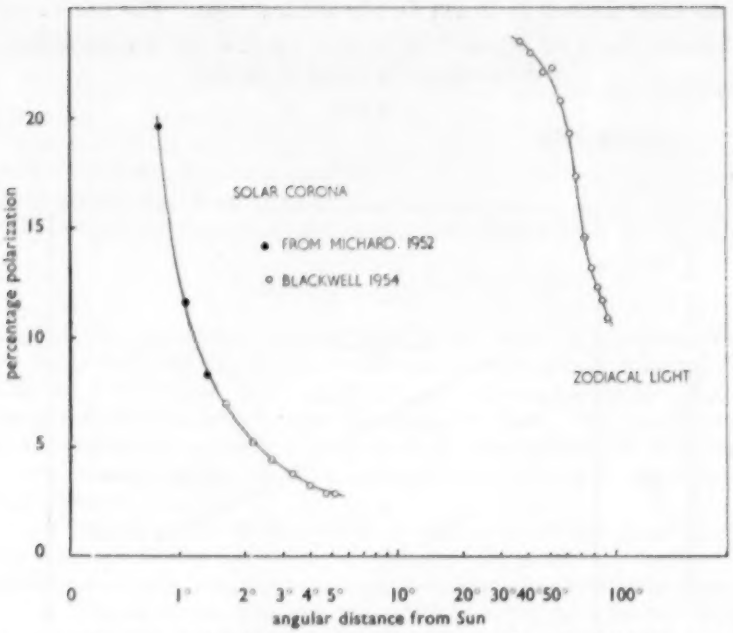


FIG. 11.—*Polarization of the solar corona and zodiacal light. The zodiacal light data are taken from the work of Behr and Siedentopf.*

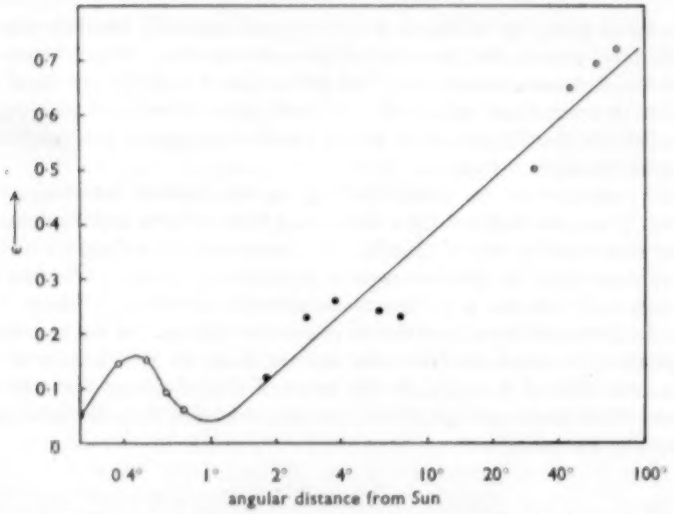


FIG. 12.—*The Ludendorff index s. Values for the inner corona are taken from the work of van de Hulst. Those for the outer corona are the author's measurements, and for the zodiacal light the isophotes of Behr and Siedentopf have been used.*

calculated using the isophotes given by Behr and Siedentopf (*loc. cit.*) slightly extrapolated to the poles.

The increase in the flattening from the limb to a maximum at a distance of about 0°·4 is due to the increasing predominance of the equatorial K corona. Dilution of the K corona by the F corona, which is symmetrical around the Sun at small distances, causes the decrease to a minimum at about 1°. Thereafter the flattening increases to the large value characteristic of the zodiacal light. It is satisfactory that the isophotes observed show a plausible value for ϵ.

The interpretation of the brightness and polarization curves will be considered in a later paper.

Acknowledgments.—It is a pleasure to acknowledge the inspiration of Professor R. O. Redman who first suggested to the author the desirability of making astronomical observations from very high altitudes.

We also acknowledge with gratitude the active cooperation of the Air Ministry who not only provided the special Lincoln aircraft but also made arrangements for testing the two Cambridge observers in a decompression chamber. The aircraft was operated from R.A.F. Upwood, and we wish to make particular mention of the great efforts, cheerfully made by the Commanding Officer, Gp. Capt. R. A. C. Carter, and all his staff, to make the venture a success.

The eclipse camera was made in the Observatories' workshops by Mr J. P. Hignell, who also helped with all the preliminary test flights and took part in the eclipse observations. The camera mounting was designed and made at R.A.F. Upwood.

The colour temperature and emissivity at various wave-lengths of the standard lamp were measured by the National Physical Laboratory.

Funds for the work, except for the running costs of the aircraft, were provided by the Joint Permanent Eclipse Committee of the Royal Society.

The Observatories,
Cambridge:
1955 *October* 20

References

(1) H. C. van de Hulst, *The Solar System*, Vol. 1, ed. G. P. Kuiper, p. 261 et seq. 1953.
(2) S. Baumbach, *Astr. Nach.*, **263**, 121, 1937.
(3) R. Michard, A. Dollfus, J.-C. Pecker, M. Laffineur, M. d'Azambuja, *Ann. d'Astrophys.*, **17**, 320, 1954.
(4) W. A. Rense, J. M. Jackson, B. Todd, *Journ. Geophys. Res.*, **58**, 369, 1953.
(5) C. W. Allen, *M.N.*, **107**, 426, 1947.
(6) M. Minnaert, *B.A.N.*, **2**, 75, 1924.
(7) E. Schoenberg, *Handbuch der Astrophysik*, Vol. **2**/1, p. 264, 1929.
(8) Y. Öhman, *Stock. Obs. Ann.*, **15**, No. 2, 1947.
(9) H. C. van de Hulst, *B.A.N.*, **11**, 135, 1950.
(10) C. W. Allen, *M.N.*, **101**, 281, 1941.
(11) F. E. Roach, H. B. Pettit, E. Tandberg-Hanssen, D. N. Davis, *Ap. J.*, **119**, 253, 1954.
(12) A. Behr, H. Siedentopf, *Zeits. f. Astrophys.*, **32**, 19, 1953.

A VARIATIONAL METHOD FOR IMPROVING MODEL STELLAR ATMOSPHERES

Antoni Przybylski

(Communicated by the Commonwealth Astronomer)

(Received 1955 September 10)

Summary

A variational method of improving model stellar atmospheres has been developed and applied to a model solar atmosphere with a net flux increasing by 35 per cent from the top of the atmosphere to the mean optical depth $\tau=12$. The net flux of the improved model is constant within 1 per cent above the mean optical depth $\tau=1{\cdot}7$ and within 4·4 per cent above the mean optical depth $\tau=11{\cdot}5$.

The use of the Rosseland mean absorption coefficient and Milne's law of the temperature distribution leads to good first approximations in the computation of model atmospheres for cool stars.

---

1. *Introduction.*—In recent investigations on model stellar atmospheres several authors (**1**–**5**) computed the net flux of energy at different optical depths in order to check its constancy. The results were generally disappointing, especially for hotter models, and this stimulated further investigations on the possibility of improving the models so that constancy of the net flux was obtained. An iterative method proposed by B. Strömgren, and based on the assumption that at every level the absorption must be equal to the emission, was recently applied in practice to a hot model by J. C. Pecker (**2**), who, however, even after six iterations failed to achieve the required constancy of net flux. Although this method may possibly still give good results for cooler model stellar atmospheres, the present paper develops a variational method and applies it to a model solar atmosphere.

2. *Model solar atmosphere: first approximation.*—A model solar atmosphere computed in continuation of the author's investigations concerning the effect of convection in stellar atmospheres (**6, 7**) was used as the initial basis for the present investigations. This model was not included in the previous investigations on stellar atmospheres.

The Planck mean coefficient of absorption was used in the computation of this model for reasons explained in the first paper (**6**) dealing with this subject, although in the present case to a certain extent the use of the Rosseland mean as based on the theory of radiative equilibrium would be more advantageous. Since, however, the primary object of the present investigation is to test the usefulness of a method of improving model stellar atmospheres, the choice of the mean value of the absorbtion coefficient for the construction of the initial model is of minor importance.

The computations of the solar model followed closely the method described in a previous paper (**6**). The assumed value of $\log A$ is 3·8, where A is the ratio of the

number of hydrogen particles to the number of metal particles. The model is given in Table I. Since helium was disregarded in the computations, p in Table I means the total pressure of hydrogen alone.

TABLE I

Model solar atmosphere. First approximation

τ	θ	$\log_{10}p$	$\log_{10}p_e$
0·00	1·0439		
0·01	1·040	3·89	0·03
0·02	1·036	4·05	0·17
0·03	1·032	4·14	0·26
0·04	1·029	4·21	0·33
0·05	1·025	4·26	0·38
0·06	1·022	4·30	0·42
0·07	1·018	4·34	0·45
0·08	1·015	4·37	0·48
0·09	1·011	4·40	0·51
0·10	1·008	4·42	0·54
0·12	1·002	4·47	0·58
0·14	0·995	4·50	0·62
0·16	0·989	4·53	0·66
0·18	0·983	4·56	0·69
0·20	0·978	4·59	0·72
0·25	0·964	4·64	0·79
0·30	0·951	4·68	0·85
0·40	0·928	4·75	0·95
0·50	0·908	4·80	1·05
0·60	0·889	4·84	1·13
0·70	0·872	4·88	1·21
0·80	0·857	4·90	1·30
0·90	0·843	4·93	1·38
1·00	0·830	4·95	1·46
1·047	0·824	4·95	1·49

τ	Radiative model			Adiabatic model		
	θ	$\log_{10}p$	$\log_{10}p_e$	θ	$\log_{10}p$	$\log_{10}p_e$
1·2	0·807	4·98	1·60	0·808	4·98	1·60
1·4	0·787	5·00	1·74	0·792	5·00	1·71
1·6	0·769	5·02	1·87	0·779	5·02	1·80
1·8	0·753	5·03	1·98	0·769	5·03	1·88
2·0	0·738	5·04	2·09	0·759	5·05	1·94
2·5	0·707	5·06	2·32	0·743	5·07	2·07
3·0	0·682	5·07	2·51	0·731	5·09	2·17
4·0	0·642	5·09	2·82	0·713	5·13	2·32
5·0	0·611	5·10	3·05	0·700	5·15	2·42
6·0	0·587	5·10	3·24	0·690	5·17	2·50
7·0	0·566	5·11	3·40	0·681	5·18	2·57
8·0	0·550	5·11	3·52	0·675	5·20	2·63
9·0	0·535	5·11	3·63	0·669	5·21	2·68
10·0	0·522	5·11	3·73	0·664	5·22	2·72
15·0	0·474	5·12	4·09	0·645	5·26	2·88
20·0	0·442	5·12	4·30	0·632	5·29	2·99
25·0	0·419	5·12	4·45	0·623	5·31	3·07
30·0	0·401	5·12	4·55	0·616	5·33	3·14

As in the former papers of the author (**6, 7**) two models were computed for the convectively unstable zone using the radiative and convective temperature gradients respectively in the computations. No use, however, will be made of the convective model in the present paper.

2.1. *Computation of the net flux.*—As the preparatory step for the computation of monochromatic fluxes the monochromatic depths τ_ν were found as functions of the reciprocal temperature θ for the following 28 wave-lengths in A:

1100+, 1242, 1424, 1674, 2030, 2512−, 2512+, 2975, 3647−, 3647+, 4235, 5050, 6252, 8206−, 8206+, 9214, 10503, 12213, 14588−, 14588+, 16031, 17790, 19984, 22794−, 22794+, 30392, 45592, 91185.

This procedure has already been described elsewhere (**6**).

Although the mean absorption coefficient used in the construction of the solar model was computed for neutral hydrogen and its negative ion only, the absorption by metals as well as Thomson and Rayleigh scattering were taken into account in the computation of the monochromatic depths. For this reason the values of the

absorption coefficient published by E. Vitense (**9**) were used in the ultra-violet; although these values were computed for an atmosphere for which $\log A = 3{\cdot}7$ (differing slightly from the value $\log A = 3{\cdot}8$ which was used in the computation of the solar model), the uncertainty in the contribution by metals to the absorption is so great that a recomputation of the absorption in the ultraviolet would hardly be justified. The addition of the contribution by metals and other sources of opacity to the absorption coefficient at this stage of computation is to a certain extent inconsistent with their omission in the computation of the initial model. This is, however, of minor importance. The essential point of the present paper is that a method of improving model stellar atmospheres was applied to a model given in Table I, irrespective of the way in which it was computed, and that all the known sources of absorption were taken into account in these investigations.

The monochromatic fluxes $F_\nu^{(1)}(\tau_\nu)$ were computed by numerical evaluation of the integrals

$$F_\nu^{(1)}(0) = 2\pi \int_0^\infty B_\nu(t_\nu)\, E_2(t_\nu)\, dt_\nu \quad \text{for} \quad \tau_\nu = 0,$$

$$F_\nu^{(1)}(\tau_\nu) = 2\pi \int_{\tau_\nu}^\infty B_\nu(t_\nu)\, E_2(t_\nu - \tau_\nu)\, dt_\nu - 2\pi \int_0^{\tau_\nu} B_\nu(t_\nu)\, E_2(\tau_\nu - t_\nu)\, dt_\nu \quad \text{for} \quad \tau_\nu \neq 0.$$

(The superscript (1) in these formulae means that the flux is computed for the *first* approximation of the solar model.)

Usually Cotes' formulae were used for the evaluation of the monochromatic fluxes but some use was also made of tables and formulae of A. Reiz (**8**) when they became available in the course of computations. The monochromatic fluxes were computed for $\tau_\nu = 0{\cdot}0$, $0{\cdot}2$, $0{\cdot}4$, $0{\cdot}8$, $2{\cdot}0$, $4{\cdot}0$, $8{\cdot}0$, $12{\cdot}0$ and for other values of τ_ν when necessary. The computed values of $F_\nu^{(1)}$ were then plotted against the *mean* optical depth τ, corresponding to the monochromatic depth τ_ν for which the flux was computed, and the curves of $F_\nu^{(1)}(\tau)$ were drawn for all the 28 wavelengths in question. For a number of suitable values of τ the values of $F_\nu^{(1)}(\tau)$ were taken from this diagram and finally the net flux $F^{(1)}(\tau)$ was computed by numerical evaluation of the integral

$$F^{(1)}(\tau) = \int_{\nu_0}^\infty F_\nu^{(1)}(\tau)\, d\nu,$$

where ν_0 is the frequency corresponding to $\lambda = 1100$ A. The whole radiation beyond $\lambda = 1100$ A was neglected, since there the flux is certainly very small. This is due to the absorption of radiation by carbon, which becomes very strong beyond 1100 A.

The net flux of radiation was computed for 13 values of τ. The results are listed in Table II ; the last column of this table will be explained in Section 2.3.

The net flux is reasonably constant down to the mean optical depth $\tau = 2{\cdot}6$, but at greater depths it increases and is about 35 per cent too high at the optical depth $\tau = 12$.

2.2. *Improved values of the mean absorption coefficient.*—Although the net flux varies with the optical depth the results obtained in the preceding section were used to compute the second approximation of the mean absorption coefficient with the formula

$$\kappa^{(1)} = \frac{1}{F^{(1)}} \int_0^\infty \kappa_\nu F_\nu^{(1)}\, d\nu \qquad (1)$$

TABLE II

Net flux of the first approximation in units of 10^5 *erg* cm^{-2} sec^{-1}

τ	θ	$F^{(1)}(\tau)$	corrected τ^*
0·0	1·0439	62·48	0·0
0·0101	1·04	62·35	0·0102
0·1250	1·00	61·60	0·1242
0·3053	0·95	61·36	0·2981
0·5400	0·90	61·77	0·5232
0·8499	0·85	62·40	0·8210
1·266	0·80	63·27	1·2195
1·835	0·75	63·75	1·757
2·631	0·70	64·63	2·511
3·768	0·65	66·67	3·581
5·442	0·60	70·49	5·080
7·985	0·55	77·85	7·177
12·000	0·50	84·67	10·148

which should give the correct value of κ if the net flux is constant. Since it was intended to include all the sources of absorption in the further investigations, the values of κ_ν published by E. Vitense (**9**) were used in the formula (1). The results of these evaluations of the mean coefficient for seven values of θ are given in the third column of Table III. For comparison the values of the Rosseland, Planck and Chandrasekhar means are given in additional columns; the first was taken from E. Vitense's publication (**9**), the second from the author's paper (**6**) and the last from the work of G. Münch on a model solar atmosphere (**10**). All the values are referred to one heavy particle as defined by E. Vitense (**9**), corresponding to 0·8437 particles of hydrogen. A comparison of all the values shows that for lower temperatures the numerical values agree fairly well, but for higher temperatures only the Rosseland mean follows the second approximation closely. The values of the Planck and Chandrasekhar means in Table III do not take into account metals and other sources of absorption besides neutral hydrogen and its negative ion.

TABLE III

The quantities tabulated are the values of κ/p_e *in the units of* 10^{-24} cm^2 dyn^{-1} *referred to one heavy particle as defined by E. Vitense* (**9**)

θ	$\log_{10}p$	$\kappa^{(2)}/p_e$	Rosseland mean	Planck mean	Chandrasekhar mean	$\frac{\partial\kappa/p_e}{\partial \log_{10}p_e}$
1·04	0·029	8·935	8·995	8·798	8·265	−0·026
1·00	0·591	7·365	7·499	7·564	7·024	−0·024
0·90	1·081	4·851	5·070	5·060	4·641	−0·035
0·80	1·650	3·126	3·273	3·297	2·961	−0·075
0·70	2·373	2·088	2·138	2·196		−0·135
0·60	3·137	1·528	1·567	1·763		−0·360
0·50	3·897	1·433	1·496	2·100		−0·635

It should be stressed that the computed values of the mean absorption coefficient are valid only for the actual values of electron pressure p_e at the optical depth with the reciprocal temperature θ. For this reason the values of $\log_{10} p_e$ are given together with the values of θ and $\kappa^{(2)}/p_e$ in Table III. If the values of p_e differ only slightly from those given in Table III the values of $\kappa^{(2)}/p_e$ can be computed with the aid of the differential formula

$$\frac{\kappa(\theta, p_e + \Delta p_e)}{p_e + \Delta p_e} = \frac{\kappa(\theta, p_e)}{p} + \frac{\partial(\kappa/p_e)}{\partial \log_{10} p_e} \cdot \Delta \log_{10} p_e \tag{2}$$

The present investigations are not sufficient to compute the derivative

$$\frac{\partial(\kappa/p_e)}{\partial \log_{10} p_e}.$$

Since, however, the second approximation does not differ much from the Rosseland mean, it can be assumed that this derivative for the second approximation is equal to the corresponding derivative for the Rosseland mean, which can be found easily from the data in E. Vitense's paper (**9**). The last column of Table III gives this derivative as determined graphically.

2.3. *Corrected values of the optical depth.*—The mean optical depth in the first approximation of the solar model was computed with the Planck mean which differs from the corrected value, especially at great optical depth. Since, however, for further investigations it is important that the optical depth be measured in terms of the true mean absorption coefficient, the mean optical depth was recomputed by using the second approximation of the mean absorption coefficient. A graph of $\kappa^{(2)}/p_e$ (from the third column of Table III) against θ was used to find the absorption coefficient for intermediate values of θ; the same graph was used later for the computation of the second approximation of the solar model. The new values of the optical depth for which the notation τ^* will be used were found with the formula

$$\tau^* = \frac{1{\cdot}1853}{g\, m_H} \int_0^p \kappa^{(2)}\, dp = \frac{1{\cdot}1853}{g\, m_H} \int_0^p \left(\frac{\kappa^{(2)}}{p_e}\right) \cdot p_e\, dp \tag{3}$$

where g is solar gravity, m_H the mass of the hydrogen atom and $\kappa^{(2)}$ is the new value of the absorption coefficient from the preceding section. The factor 1·1853 arises from the conversion of the mean absorption coefficient per one heavy particle to its value referred to one hydrogen atom.

The new values of τ^* are given in the last column of Table II for the levels at which the net flux of energy was computed.

3. *Variational method of improving stellar models*

3.1. *Theoretical remarks.*—The reason why the net flux is not constant is that the temperature distribution in the model is not correct. Therefore, by improving the temperature distribution one can also improve the constancy of the net flux.

Suppose that at the level τ_ν the actual emission $B_\nu(\tau_\nu)$ is erroneous and that $B_\nu(\tau_\nu) + \Delta B_\nu(\tau_\nu)$ is its correct value. If $B_\nu(\tau_\nu)$ is corrected at every level then the net flux would become constant, this net flux should be computed in the way described in Section 2.1. Introducing the operator Φ_x defined by the equation

$$\Phi_x\{f(t)\} = 2\int_x^\infty f(t)\, E_2(t-x)\, dt - 2\int_0^x f(t)\, E_2(x-t)\, dt$$

one finds the following equation for the net flux $F^{(2)}$ of the second approximation

at the mean optical level τ_0:

$$F^{(2)}(\tau_0) = \int_0^\infty F_\nu^{(2)}(\tau_0)\, d\nu = \pi \int_0^\infty \Phi_{\tau_{\nu,0}} \{B_\nu(\tau_\nu) + \Delta B_\nu(\tau_\nu)\}\, d\nu,$$

where $\tau_{\nu,0}$ is the monochromatic depth corresponding to the mean depth τ_0. Further we get

$$F^{(2)}(\tau_0) = \pi \int_0^\infty \Phi_{\tau_{\nu,0}} \{B_\nu(\tau_\nu)\} d\nu + \pi \int_0^\infty \Phi_{\tau_{\nu,0}} \{\Delta B_\nu(\tau_\nu)\}\, d\nu. \tag{4}$$

The first integral on the right-hand side is known, since it is the net flux $F^{(1)}(\tau_0)$ of the stellar model in the first approximation. The second integral is the correction which should be applied to the flux $F^{(1)}(\tau_0)$ in order to make the flux constant; it will be denoted here by $\Delta F^{(1)}(\tau_0)$.

Suppose now that the computation of the second integral can be simplified by reducing the general case in which the absorption coefficient κ_ν is a function of frequency to the grey case where it is independent of the frequency. In this case one can write

$$\Delta F^{(1)}(\tau_0) = \pi \Phi_{\tau_0} \{\Delta B(\tau)\},$$

where $\Delta B(\tau)$ is the correction to be applied to the total emission at the level τ. Suppose further that $\Delta B(\tau)$ can be represented by a linear combination of some elementary functions $f_1, f_2, \ldots, f_n$, for which the Φ-transforms are known:

$$\Delta B(\tau) = a_1 f_1(\tau) + a_2 f_2(\tau) + \ldots + a_n f_n(\tau).$$

Then

$$\Delta F^{(1)}(\tau_0) = a_1 \pi \Phi_{\tau_0}\{f_1(\tau)\} + a_2 \pi \Phi_{\tau_0}\{f_2(\tau)\} + \ldots + a_n \pi \Phi_{\tau_0}\{f_n(\tau)\}. \tag{5}$$

The differences $\Delta F^{(1)}(\tau_0)$ between the required constant flux $F^{(2)}$ and the actual flux $F^{(1)}(\tau_0)$ are of course known. Therefore, the coefficients $a_1, a_2, \ldots a_n$ are the only unknowns in equation (5). This equation can be written for all the values τ_0 for which the net flux was computed in the first approximation and then this system of equations can be solved for $a_1, a_2, \ldots, a_n$. The method of least squares should be applied if the number of equations exceeds the number of parameters. When this is done the corrected temperature $T^{(2)}(\tau)$ is to be found from the equation

$$\left.\begin{aligned} \frac{\sigma}{\pi}\{T^{(2)}(\tau)\}^4 &= B(\tau) + \Delta B(\tau) = B(\tau) + a_1 f_1(\tau) + a_2 f_2(\tau) + \ldots + a_n f_n(\tau) \\ &= \frac{\sigma}{\pi}\{T^{(1)}(\tau)\}^4 + a_1 f_1(\tau) + a_2 f_2(\tau) + \ldots + a_n f_n(\tau) \end{aligned}\right\} \tag{6}$$

where $T^{(1)}(\tau)$ is the temperature at the level τ in the first approximation of the stellar model and σ is the Stefan–Boltzmann constant. In this way the corrected law of the temperature distribution in the stellar atmosphere is found.

The method described above is likely to lead to good results if the corrections to the net flux are small and if the mean optical depth is measured in terms of the true mean absorption coefficient for both the initial and the corrected model. For this reason the optical depth τ of the initial model was recomputed in Section 2.3. In all the computations τ should be taken from the last column of Table II and therefore the notation τ^* will be used in the next section.

3.2. *Application to the solar model.*—The method described in the preceding section was applied to the solar model given in Table I. The corrections $\Delta B(\tau^*)$ were represented by a linear combination of the functions

$$f_1 = 1,\ f_2 = \tau^*,\ f_3 = \tau^{*2},\ f_4 = \tau^{*3},\ f_5 = e^{-\tau^*},$$

for which the Φ-transforms are $2E_3(\tau^*)$, $\frac{1}{3} - 2E_4(\tau^*)$, $\frac{\pi}{8}\tau^* + 4E_5(\tau^*)$, $4{\cdot}8 - 4\tau^{*2} - 12E_6(\tau^*)$ and $2\{e^{-\tau^*}(0{\cdot}72963 - \ln\tau^*) - E_1(\tau^*) - E_2(\tau^*)\}$ respectively.

The system of 13 equations (5) with five unknowns a_1, a_2, a_3, a_4 and a_5 was solved by the least-squares method. It was intended to make the net flux equal to the boundary value $62{\cdot}48 \times 10^9$ erg cm^{-2} sec^{-1} although this value is a little lower than the most probable value of the solar flux. More weight was given to the equations with low value of τ^* in order to make the flux more constant in the outer layers. This, however, led to a big residual of $1{\cdot}43 \times 10^9$ erg cm^{-2} sec^{-1} in the net flux for $\tau^* = 7{\cdot}177$. Further investigations (Section 4.2) show that even larger errors are possible for intermediate values of τ^* not shown in Table II and therefore not included in the least-square solution; indeed, Table VI shows that for $\tau^* = 7{\cdot}740$ one can expect a residual of $2{\cdot}92 \times 10^9$ erg cm^{-2} sec^{-1} if this level would be taken into account in the least-square solution. By giving equal weight to all the equations one could reduce the residual for greater optical depths, but in this case the residual for smaller values of τ^* would be increased.

The following values of the parameters a were obtained in units of 10^9 erg cm^{-2} sec^{-1}:

$a_1 = -0{\cdot}03785$, $a_2 = +0{\cdot}64892$, $a_3 = -0{\cdot}22394$, $a_4 = -0{\cdot}00481$, $a_5 = -0{\cdot}09937$.

With these values for the parameters the new corrected values of the temperature were found by means of the formula (6) for the 13 levels at which the flux was computed. They are listed in Table IV as $T^{(2)}$ together with the previous values $T^{(1)}$.

TABLE IV

The temperature distribution in the first and second approximations

τ^*	$T^{(1)}$	$T^{(2)}$	Deviation from Milne's law $T^4 = \frac{1}{2} T_e^4 (1 + \frac{3}{2}\tau^*)$
0·0	4828°K	4811·0°K	−33·4°
0·0102	4846	4830·3	−32·5
0·1242	5040	5034·7	−21·1
0·2981	5305	5311·0	− 2·4
0·5232	5600	5614·2	+14·9
0·8210	5929	5949·1	+28·2
1·2195	6300	6321·1	+38·2
1·757	6720	6736·7	+47·5
2·511	7200	7203·5	+45·5
3·581	7754	7729·6	+32·9
5·080	8400	8325·4	+24·6
7·177	9164	9003·0	+30·9
10·148	10080	9775·3	+53·1

It can easily be shown that the new values of the temperature differ but little from the values given by Milne's law

$$T^4 = \tfrac{1}{2} T_e^4 (1 + \tfrac{3}{2}\tau) \qquad (7)$$

for a model with the effective temperature $T_e = 5761$ deg.K corresponding to a net flux equal to $62{\cdot}48$ erg cm^{-2} sec^{-1}. These deviations from Milne's law are given in the last column of Table IV.

4. *Solar model atmosphere: second approximation.*—In the previous sections were found the new corrected values $\kappa^{(2)}$ of the mean absorption coefficient and the

second approximation $T^{(2)}$ to the temperature as a function of the optical depth. It remains now to construct a model atmosphere in the second approximation, i.e. to calculate the pressure as a function of the optical depth using the new values of the mean absorption coefficient and the new law of the temperature distribution; and finally to compute the net flux at various optical depths, since the degree of constancy of the flux in the second approximation indicates how successful the second approximation is.

The notation τ will be used for the mean optical depth. This optical depth τ of the second approximation corresponds to the optical depth τ^* of the first approximation, since both are measured in terms of the same mean absorption coefficient.

4.1. *Computation of the model.*—The following auxiliary graphs were constructed for the calculation of the second approximation:—

1. A graph giving the relation between $\kappa^{(2)}/p_e$ and θ (graph 1); this graph was already used for the construction of the improved model in Section 2.3;

2. A graph giving the relation between $\partial(\kappa/p_e)/\partial \log_{10} p_e$ and θ (graph 2); this graph was prepared from the data in the last column of Table III;

3. A graph giving the relation between $\log p_e$ and θ for the first approximation of the solar model (graph 3);

4. A graph giving the correction of Milne's law of temperature distribution (graph 4); this graph with the argument τ was prepared from the data in Table IV.

The computation of the new model consisted in the evaluation of the integral in equation (3) in which τ^* is substituted by τ. The numerical procedure was described by the author in a former paper (**6**). In each step of integration τ was first extrapolated and then with Milne's equation (7) and graph 4 the corrected temperature was found. Next the electron pressure was computed and compared with the electron pressure for the initial model found from graph 3 for the same value of θ. Then $\kappa^{(2)}/p_e$ was found from graph 1 and, if the electron pressure given by the second approximation differed from that given by the first, corrected with the aid of formula (2). The value of $\partial(\kappa^{(2)}/p_e)/\partial \log_{10} p_e$ for this correction was taken from graph 2. When the corrected value of $\kappa^{(2)}/p_e$ was found, the integrand in equation (3) was determined and τ was computed. If the value of τ found in that way differed from the extrapolated one, the whole process of integration was repeated.

It should be stressed that the corrections of $\kappa^{(2)}/p_e$, which were found by means of formula (2), were always small and in no case exceeded 0·3 per cent of the value of $\kappa^{(2)}/p_e$; this means that from the practical point of view they could be disregarded altogether. In other stellar models this correction may, however, be more important.

The solar model computed in this way is shown in Table V.

Graph 1 gives the values of $\kappa^{(2)}/p_e$ between $\theta = 0{\cdot}5$ and 1·04. Beyond $\theta = 1{\cdot}04$ the values of $\kappa^{(2)}/p_e$ were extrapolated graphically and for $\theta < 0{\cdot}5$ the Rosseland mean, taken from E. Vitense's paper (**9**), was used.

4.2. *Net flux.*—The net flux was computed in the same way as in the initial model. The only difference was that the monochromatic fluxes were plotted against the reciprocal temperature θ instead of τ, which proved to be more advantageous in practice. The results are shown in Table VI.

The net flux is fairly constant throughout the whole atmosphere; the greatest deviation from the arithmetical mean of 13 values ($62{\cdot}64 \times 10^9$ erg cm^{-2} sec^{-1}) is 4·4 per cent for $\tau = 7{\cdot}74$. The boundary value of the flux does not differ much

from the mean value for the whole atmosphere; it is only 0·3 per cent higher than the previous value, which was supposed to be kept constant. The boundary value of the flux corresponds to a solar constant 1·945 cal cm^{-2} min^{-1}, which is certainly not far from the actual value.

TABLE V

Model solar atmosphere. Second approximation.

τ	θ	$\log_{10}p$	$\log_{10}p_e$
0·0	1·0476		
0·01	1·043	3·88	0·02
0·02	1·039	4·04	0·16
0·03	1·035	4·14	0·25
0·04	1·032	4·21	0·32
0·05	1·028	4·26	0·37
0·06	1·024	4·30	0·41
0·07	1·020	4·34	0·45
0·08	1·017	4·37	0·48
0·09	1·013	4·40	0·51
0·10	1·009	4·42	0·54
0·12	1·002	4·47	0·58
0·14	0·996	4·50	0·62
0·16	0·989	4·54	0·66
0·18	0·983	4·57	0·70
0·20	0·977	4·59	0·73
0·25	0·962	4·64	0·80
0·30	0·948	4·69	0·86
0·35	0·936	4·72	0·91
0·40	0·924	4·76	0·97
0·45	0·913	4·78	1·02
0·50	0·902	4·81	1·06
0·60	0·883	4·85	1·15
0·70	0·866	4·88	1·24
0·80	0·850	4·91	1·33
0·90	0·836	4·93	1·42
1·0	0·823	4·95	1·50
1·2	0·799	4·98	1·65
1·4	0·779	5·00	1·79
1·6	0·761	5·02	1·92
1·8	0·745	5·03	2·04
2·0	0·730	5·04	2·14
2·5	0·700	5·06	2·37
3·0	0·676	5·07	2·56
3·5	0·655	5·08	2·71
4·0	0·637	5·09	2·85
4·5	0·621	5·09	2·97
5·0	0·607	5·10	3·08
6·0	0·583	5·10	3·26
7·0	0·563	5·11	3·42
8·0	0·546	5·11	3·55
9·0	0·531	5·11	3·67
10·0	0·517	5·11	3·77
15·0	0·470	5·12	4·12
20·0	0·438	5·12	4·33
25·0	0·415	5·12	4·48
30·0	0·397	5·13	4·57

TABLE VI

Net flux for the second approximation in units of 10^9 *ergs cm*$^{-1}$ *sec*$^{-1}$

τ	θ	$F^{(1)}(\tau)$
0·0	1·0476	62·68
0·0186	1·04	62·51
0·1274	1·00	62·05
0·2943	0·95	62·03
0·5116	0·90	62·32
0·8021	0·85	62·63
1·1950	0·80	62·92
1·733	0·75	62·70
2·505	0·70	61·89
3·636	0·65	61·38
5·291	0·60	62·35
7·740	0·55	65·40
11·528	0·50	63·51

5. *Convection.*—In the computations of the second approximation of the solar model it was assumed that there is radiative equilibrium throughout the whole atmosphere, since otherwise the computation of the mean absorption coefficient with the formula (1) would not be justified. It should be stressed, however, that the atmosphere becomes convectively unstable at the optical depth $\tau = 1{\cdot}014$. No attempt was made to take convection into account.

6. *Blanketing effect.*—The darkening of the limb was computed for a number of wavelengths for the solar model in the first approximation and then compared with the observational data. Generally the agreement between the computed and observed darkening of the limb was not very good, especially far from the centre of the disk. Some differences may be due to convection, but probably the main source of these discrepancies is the blanketing effect. It is intended to include the blanketing effect in further investigations.

7. *Comparison of the model solar atmosphere with observation.*—Although the blanketing effect considerably modifies the darkening of the solar limb, it has little effect on some other properties of the solar spectrum. Therefore, in spite of the total omission of the blanketing effect, some results derived from the theoretical model can be compared with the observational data.

It is not intended to make a detailed comparison between the model solar atmosphere and the observations in this paper. Only two features of our model are briefly mentioned below.

7.1. *Solar constant.*—The computed emerging net flux of our model corresponds to a solar constant $1{\cdot}945\,\mathrm{cal\,cm^{-2}\,min^{-1}}$. This value does not differ much from the actual value of the solar constant, which according to Nicolet (**11**) is $1{\cdot}98\,\mathrm{cal\,cm^{-2}\,min^{-1}}$, and agrees well with Abbot's value of $1{\cdot}94\,\mathrm{cal\,cm^{-2}\,min^{-1}}$.

7.2. *Balmer discontinuity.*—Our solar model shows a discontinuity at the Balmer limit

$$D = \log_{10}(I_0^{+}/I_0^{-}) = 0{\cdot}114,$$

where I_0^{+} is the intensity of radiation for the centre of the solar disk for $\lambda = 3647+$ and I_0^{-} the same intensity for $\lambda = 3647-$. This value of D does not differ much from the observed value $D = 0{\cdot}125$ (**12**).

8. *Final remarks.*—The differences between the values of the net flux of the second approximation $F^{(2)}(\tau)$ and its boundary value $F^{(2)}(0)$ are not scattered at random but show a definite systematic run with the optical depth. This means that the second approximation of the solar model should not be considered as the final one, since the repetition of the whole procedure can still improve the constancy of the net flux. The second approximation shows, however, that the method described in Section 3.1 works well. This is due probably to the following three factors:

(i) The operator $\mathbf{\Phi}_x$ used in Section 3.1 simultaneously takes into account the corrections $\Delta B(\tau)$ for all the values of τ.

(ii) The corrected values of the mean absorption coefficient are used in the computations.

(iii) The mean value of the absorption coefficient $\kappa^{(2)}$ is used essentially only in the variational corrections

$$\Delta F^{(1)}(\tau_0) = \pi \mathbf{\Phi}_{\tau_0}\{\Delta B(\tau)\},$$

which in Section 3.1 are used in order to find the new law of the temperature distribution. No appeal to a mean absorption coefficient is made in the integral

$$\pi \int_0^\infty \Phi_{\tau_{\nu,0}}\{B_\nu(\tau_\nu)\}\, d\nu,$$

which gives the main contribution to the net flux $F^{(2)}(\tau_0)$ in equation (4). This integral was not computed explicitly in the second approximation for the stellar model, but it enters correctly into the theoretical consideration of this approximation, because the same mean absorption coefficient was used in the initial and the corrected models. It was in order to bring this about that the values of τ in the initial model were re-computed with $\kappa^{(2)}$ in Section 2.3.

Table IV shows that Milne's law (7) is a good approximation to the temperature distribution, if the correct values of the mean absorption coefficient are used in the computation of the stellar atmosphere. On the other hand, Table III shows that the Rosseland mean does not differ by more than 5 per cent from the second approximation $\kappa^{(2)}$ of the mean absorption coefficient. The use of Milne's law of the temperature distribution and of the Rosseland mean absorption coefficient in the computation of model stellar atmospheres should therefore lead to good initial models of cool stellar atmospheres.

9. *Acknowledgment.*—The present investigations are due to the initiative of Professor R. v. d. R. Woolley, to whom the author is grateful for many valuable discussions and much advice.

Commonwealth Observatory,
Canberra :
1955 *August.*

References

(1) A. Underhill, *Publ. Dom. Ap. Obs.*, **8**, 357, 1951.
(2) J. C. Pecker, *Ann. d'Ap.*, **13**, 294, 319, 433, 1952.
(3) R. Michard, *Ann. d'Ap.*, **12**, 291, 1949.
(4) J. E. Milligan and L. H. Aller, *Proc. Conf. Stellar Atmospheres, Indiana*, ed. M. H. Wrubel, 1954.
(5) K. J. McDonald, *Publ. Dom. Ap. Obs.*, **9**, 269, 1953.
(6) A. Przybylski, *M.N.*, **113**, 683, 1953.
(7) A. Przybylski, *M.N.*, **114**, 406, 1954.
(8) A. Reiz, *Medd. Lunds Obs.*, Ser. **1**, Nr. 184, 1954.
(9) E. Vitense, *Z. Ap.*, **28**, 81, 1951.
(10) G. Münch, *Ap. J.*, **106**, 217, 1947.
(11) M. Nicolet, *Ann. d'Ap.*, **14**, 249, 1951.
(12) M. Minnaert, *The Sun.*, ch. 3, ed. G. Kuiper, Chicago, 1953.

A MATHEMATICAL VERIFICATION OF THE PRINCIPLE OF INVARIANCE AS APPLIED TO COMPLETELY NON-COHERENT SCATTERING AND TO INTERLOCKED MULTIPLET LINES

I. W. Busbridge

(Received 1955 May 26)

Summary

In some recent papers, the solutions of some problems of line formation involving non-coherent scattering and the interlocking of multiplet lines have been based on the principle of invariance governing the law of diffuse reflection. Hitherto this principle had been applied only in cases of coherent scattering and its applicability to problems involving non-coherent scattering processes has been questioned. In this paper the solutions for completely non-coherent scattering and for interlocked multiplet lines as given in (**3**) and (**5**) are obtained by a mathematical method, thus providing an independent verification of the principle of invariance in these cases. The problem of non-coherent scattering solved here is more general than that given in (**3**). It allows for the possibility of thermal emission associated with the line absorption; it applies when the frequencies involved are restricted to a finite interval (ν_1, ν_2) and it allows for incident radiation on the surface of the star.

---

1. *Introduction.*—The principle of invariance governing the law of diffuse reflection* is stated by Chandrasekhar in (**6**), p. 90, as follows:

> "The law of diffuse reflection by a semi-infinite plane-parallel atmosphere must be invariant to the addition (or subtraction) of layers of arbitrary optical thickness to (or from) the atmosphere."

This principle has been applied in the papers (**3**), (**5**), (**8**) and (**10**) and an independent verification of the solutions is desirable. This, in its turn, provides a verification of the principle of invariance in the cases considered.

Recently Sobolev has given in (**9**)† a probability method for dealing with completely non-coherent scattering, which verifies the results of (**8**) and (**3**). This method is a development of an earlier probability method of which Kourganoff wrote in (**7**):

> "Systematic analysis of Sobolev's analysis discloses that his method is entirely equivalent to that of Ambartsumian described in section 28."

The method of Ambartsumian referred to here is that given in (**2**) and I am indebted to Dr Kourganoff for lending me a translation of the paper. I soon saw that the method was far more powerful than its author can have realized. In a recent paper (**4**) I have developed it rigorously and have

* This was first formulated by Ambartsumian in (**1**) for infinitely thin layers.

† I am very much indebted to Dr J. B. Sykes for calling my attention to this paper and for translating it for me.

used it to obtain the emergent intensity from a (coherent) scattering atmosphere corresponding to various expansions of the Planck function. In this paper I use it to obtain a generalization of the solution given in (3) for non-coherent scattering. The analysis for interlocked multiplet lines is identical, except that integrals over (ν_1, ν_2) are replaced by sums, and instead of one equation holding for $\nu_1 \leqslant \nu \leqslant \nu_2$, we have k equations holding for $\nu = \nu_1, \nu_2, \ldots, \nu_k$. By introducing Stieltjes integrals or Dirac's delta function, the two problems could be dealt with simultaneously, but in order not to overload the equations, I shall give the analysis for the non-coherent case only.

The treatment here is non-rigorous, but the analysis could be justified as in (4).

2. *The equation of transfer.*—I shall make the following assumptions, which are those made in (5):

(i) The variations of the continuous absorption coefficient κ and of the Planck function $B(\nu, T)$ over the range of frequencies considered may be neglected.

(ii) The ratio η_ν of the line-absorption coefficient σ_ν to κ is independent of depth.

(iii) The Planck function is given by

$$B(\nu, T) = a + b\tau$$

where

$$\tau = \int_0^x \kappa\rho \, dx,$$

x being the depth below the surface of the atmosphere.

(iv) The coefficient ϵ, introduced to allow for thermal emission associated with the line absorption, is independent of both frequency and depth.

Subject to the conditions (i), (ii), (iii), the solution found in (3) was stated to be exact, and we shall verify the truth of that statement. On introducing ϵ and restricting the values of ν to the interval (ν_1, ν_2), the equation of transfer given in (3) becomes (see (3), equation (17))

$$\mu \frac{dI_\nu(\tau, \mu)}{d\tau} = (1 + \eta_\nu) I_\nu(\tau, \mu) - (1 + \epsilon\eta_\nu)(a + b\tau) - \tfrac{1}{2}(1 - \epsilon)\eta_\nu \int_{\nu_1}^{\nu_2} \alpha_{\nu'} \, d\nu' \int_{-1}^{1} I_{\nu'}(\tau, \mu') \, d\mu', \tag{2.1}$$

where

$$\alpha_\nu = \eta_\nu \Big/ \int_{\nu_1}^{\nu_2} \eta_{\nu'} \, d\nu', \tag{2.2}$$

so that

$$\int_{\nu_1}^{\nu_2} \alpha_\nu \, d\nu = 1. \tag{2.3}$$

In all equations in ν we have $\nu_1 \leqslant \nu \leqslant \nu_2$.

Suppose that the star is subject to incident radiation of intensity $I_\nu^{(0)}(0, -\mu')$, where $0 < \mu' \leqslant 1$. Then the equation (2.1) has to be solved subject to the condition

$$I_\nu(0, -\mu') \equiv I_\nu^{(0)}(0, -\mu') \qquad (0 < \mu' \leqslant 1, \quad \nu_1 \leqslant \nu \leqslant \nu_2). \tag{2.4}$$

Write (*cf.* (5), equation (3.8))

$$\lambda_\nu = \frac{1 + \epsilon\eta_\nu}{1 + \eta_\nu}, \qquad \eta_\nu = \frac{1}{1 + \eta_\nu}. \tag{2.5}$$

From these and (2.2) we have

$$\frac{1-\lambda_\nu}{n_\nu\alpha_\nu}=(1-\epsilon)\int_{\nu_1}^{\nu_2}\eta_{\nu'}\,d\nu'$$

and this is independent of ν. Hence we shall also write

$$\frac{1-\lambda_\nu}{n_\nu\alpha_\nu}=\varpi. \tag{2.6}$$

By (2.5), equation (2.1) can be written

$$n_\nu\mu\frac{dI_\nu(\tau,\mu)}{d\tau}=I_\nu(\tau,\mu)-\lambda_\nu(a+b\tau)-\tfrac{1}{2}(1-\lambda_\nu)\int_{\nu_1}^{\nu_2}\alpha_{\nu'}\,d\nu'\int_{-1}^{1}I_{\nu'}(\tau,\mu')\,d\mu'. \tag{2.7}$$

By writing

$$I_\nu(\tau,\mu)=a+b(\tau+n_\nu\mu)+I_\nu^*(\tau,\mu) \tag{2.8}$$

and using (2.3), we have (*cf.* (3), equations (34) and (37))

$$n_\nu\mu\frac{dI_\nu^*(\tau,\mu)}{d\tau}=I_\nu^*(\tau,\mu)-\tfrac{1}{2}(1-\lambda_\nu)\int_{\nu_1}^{\nu_2}\alpha_{\nu'}\,d\nu'\int_{-1}^{1}I_{\nu'}^*(\tau,\mu')\,d\mu', \tag{2.9}$$

and, by (2.4), this has to be solved subject to the condition

$$I_\nu^*(0,-\mu')\equiv f_\nu(\mu')\qquad(0<\mu'\leqslant 1,\quad \nu_1\leqslant\nu\leqslant\nu_2), \tag{2.10}$$

where

$$f_\nu(\mu')=I_\nu^{(0)}(0,-\mu')-a+bn_\nu\mu'. \tag{2.11}$$

As the boundary condition for large τ, we shall assume that $I_\nu(\tau,\mu)$ is at most $O(\tau)$ for larger τ. The same is then true for $I_\nu^*(\tau,\mu)$.

3. *The Milne integral equation.*—Equation (2.9) can be written

$$n_\nu\mu\,dI_\nu^*(\tau,\mu)/d\tau=I_\nu^*(\tau,\mu)-\mathfrak{S}_\nu(\tau), \tag{3.1}$$

where

$$\mathfrak{S}_\nu(\tau)=\tfrac{1}{2}(1-\lambda_\nu)\int_{\nu_1}^{\nu_2}\alpha_{\nu'}\,d\nu'\int_{-1}^{1}I_{\nu'}^*(\tau,\mu')\,d\mu'. \tag{3.2}$$

From (3.1) and the boundary conditions,

$$I_\nu^*(\tau,\mu)=e^{\tau/n_\nu\mu}\int_\tau^\infty\mathfrak{S}_\nu(t)\,e^{-t/n_\nu\mu}\frac{dt}{n_\nu\mu}\qquad(\mu>0), \tag{3.3}$$

$$I_\nu^*(\tau,\mu)=f_\nu(-\mu)\,e^{\tau/n_\nu\mu}-e^{\tau/n_\nu\mu}\int_0^\tau\mathfrak{S}_\nu(t)\,e^{-t/n_\nu\mu}\frac{dt}{n_\nu\mu}\qquad(\mu<0). \tag{3.4}$$

Hence (putting $-\mu$ for μ when $\mu<0$)

$$\int_{-1}^{1}I_\nu^*(\tau,\mu)\,d\mu=\frac{1}{n_\nu}\int_0^\infty\mathfrak{S}_\nu(t)E_1\left(\frac{|t-\tau|}{n_\nu}\right)dt+\int_0^1 f_\nu(\mu)\,e^{-\tau/n_\nu\mu}\,d\mu,$$

and (3.2) gives,

$$\mathfrak{S}_\nu(\tau)=(1-\lambda_\nu)\int_{\nu_1}^{\nu_2}\frac{\alpha_{\nu'}}{n_{\nu'}}\Lambda_{\nu'}\{\mathfrak{S}_{\nu'}(t);\,n_{\nu'}\}\,d\nu'+\tfrac{1}{2}(1-\lambda_\nu)\int_{\nu_1}^{\nu_2}\alpha_{\nu'}\,d\nu'\int_0^1 f_{\nu'}(\mu')\,e^{-\tau/n_{\nu'}\mu'}\,d\mu', \tag{3.5}$$

where

$$\Lambda_{\nu'}\{\phi(t);\,n_{\nu'}\}=\tfrac{1}{2}\int_0^\infty\phi(t)E_1\left(\frac{|t-\tau|}{n_{\nu'}}\right)dt. \tag{3.6}$$

Equation (3.5) is the Milne integral equation for $\mathfrak{S}_\nu(\tau)$.

The properties of the linear operator Λ defined by (3.6) are practically identical with those of Hopf's operator Λ (see (7), Chap. II). A slight difference occurs in the differentiation formula:

$$\frac{d}{d\tau}\Lambda_\tau\{\phi(t);\, n_{\nu'}\} = \Lambda_\tau\{\phi'(t);\, n_{\nu'}\} + \tfrac{1}{2}\phi(0)E_1\left(\frac{\tau}{n_{\nu'}}\right). \tag{3.7}$$

The operator Λ commutes with integrations with respect to variables other than τ and ν'.

If $I_\nu^*(\tau,\mu)$ is at most $O(\tau)$ for large τ, it follows from (3.1) that the same is true for $\mathfrak{I}_\nu(\tau)$. In (4) I have shown that, when this is the case, the only solution of the *homogeneous* Milne equation is $\mathfrak{I}_\nu(\tau) \equiv 0$. This is even true when $\mathfrak{I}_\nu(\tau)$ is $O(\ln \tau^{-1})$ for small τ. Moreover, I have shown that the functions appearing in the subsequent analysis by Ambartsumian's method are at most $O(\tau)$ for large τ and $O(\ln \tau^{-1})$ for small τ. We shall assume, without further proof, that the same is true in the case considered here, so that when we obtain an equation of the form

$$\phi_\nu(\tau) = (1-\lambda_\nu)\int_{\nu_1}^{\nu_0} \frac{\alpha_{\nu'}}{n_{\nu'}}\Lambda_\tau\{\phi_\nu(t);\, n_{\nu'}\}\, d\nu', \tag{3.8}$$

we can at once deduce that $\phi_\nu(\tau) \equiv 0$. Physically, this is equivalent to saying that the intensity $I_\nu^*(\tau,\mu)$ is zero throughout the atmosphere if there is no incident radiation on the surface of the atmosphere.

4. *The auxiliary equation.*—In Ambartsumian's method, everything turns on finding the appropriate auxiliary equation or equations from which the H-functions are derived. For non-coherent scattering the required equation is

$$J_\nu(\tau,\sigma_{\nu_0}) = (1-\lambda_\nu)\int_{\nu_1}^{\nu_0} \frac{\alpha_{\nu'}}{n_{\nu'}}\Lambda_\tau\{J_\nu(t,\sigma_{\nu_0});\, n_{\nu'}\}\, d\nu' + (1-\lambda_\nu)\,e^{-\tau\sigma_{\nu_0}}, \tag{4.1}$$

where $J_\nu(\tau,\sigma_{\nu_0})$ is the unknown function and $\sigma_{\nu_0} = \sigma/n_{\nu_0}$. In the analysis we shall use the following notation:

$$\sigma_{\nu_0} = \sigma/n_{\nu_0}, \qquad s_\nu = s/n_\nu, \qquad x_{\nu'} = x/n_{\nu'}, \qquad u_{\nu''} = u/n_{\nu''},$$

the letters and suffixes always occurring in these combinations.

From (4.1) it is seen that $J_\nu(\tau,\sigma_{\nu_0})/(1-\lambda_\nu)$ is independent of ν, and we therefore write

$$J_\nu(\tau,\sigma_{\nu_0}) = (1-\lambda_\nu)J(\tau,\sigma_{\nu_0}). \tag{4.2}$$

Then (4.1) becomes, by (2.6),

$$J(\tau,\sigma_{\nu_0}) = \varpi\int_{\nu_1}^{\nu_0} \alpha_{\nu'}{}^2\Lambda_\tau\{J(t,\sigma_{\nu_0});\, n_{\nu'}\}\, d\nu' + e^{-\tau\sigma_{\nu_0}}. \tag{4.3}$$

This is the final form for the auxiliary equation.

Differentiating (4.3) and using (3.7), we have

$$\frac{\partial}{\partial\tau}J(\tau,\sigma_{\nu_0}) = \varpi\int_{\nu_1}^{\nu_0} \alpha_{\nu'}{}^2\Lambda_\tau\left\{\frac{\partial}{\partial t}J(t,\sigma_{\nu_0});\, n_{\nu'}\right\} d\nu' + \tfrac{1}{2}\varpi J(0,\sigma_{\nu_0})\int_{\nu_1}^{\nu_0} \alpha_{\nu'}{}^2 E_1\left(\frac{\tau}{n_{\nu'}}\right) d\nu' - \sigma_{\nu_0}e^{-\tau\sigma_{\nu_0}},$$

and therefore

$$\frac{\partial}{\partial\tau}J(\tau,\sigma_{\nu_0}) + \sigma_{\nu_0}J(\tau,\sigma_{\nu_0}) = \varpi\int_{\nu_1}^{\nu_0} \alpha_{\nu'}{}^2\Lambda_\tau\left\{\frac{\partial}{\partial t}J(t,\sigma_{\nu_0}) + \sigma_{\nu_0}J(t,\sigma_{\nu_0});\, n_{\nu'}\right\} d\nu' + \tfrac{1}{2}\varpi J(0,\sigma_{\nu_0})\int_{\nu_1}^{\nu_0} \alpha_{\nu'}{}^2 E_1\left(\frac{\tau}{n_{\nu'}}\right) d\nu'. \tag{4.4}$$

From (4.3), since $\sigma_{\nu_0}=\sigma/n_{\nu_0}$,*

$$\int_1^\infty J(\tau,\sigma_{\nu_0})\frac{d\sigma}{\sigma}=\varpi\int_{\nu_1}^{\nu_2}\alpha_{\nu'}{}^2\Lambda_\tau\left\{\int_1^\infty J(t,\sigma_{\nu_0})\frac{d\sigma}{\sigma}\;;\;n_{\nu'}\right\}d\nu'+E_1\left(\frac{\tau}{n_{\nu_0}}\right),$$

and hence

$$\int_{\nu_1}^{\nu_2}\alpha_{\nu_0}{}^2\,d\nu_0\int_1^\infty J(\tau,\sigma_{\nu_0})\frac{d\sigma}{\sigma}=\int_{\nu_1}^{\nu_2}\alpha_{\nu_0}{}^2E_1\left(\frac{\tau}{n_{\nu_0}}\right)d\nu_0$$
$$+\varpi\int_{\nu_1}^{\nu_2}\alpha_{\nu_0}{}^2\,d\nu_0\int_{\nu_1}^{\nu_2}\alpha_{\nu'}{}^2\Lambda_\tau\left\{\int_1^\infty J(t,\sigma_{\nu_0})\frac{d\sigma}{\sigma}\;;\;n_{\nu'}\right\}d\nu'.$$

Changing the orders of integration in the last term and then changing the variables σ and ν_0, this can be written

$$\int_{\nu_1}^{\nu_2}\alpha_{\nu'}{}^2\,d\nu'\int_1^\infty J(\tau,x_{\nu'})\frac{dx}{x}=\int_{\nu_1}^{\nu_2}\alpha_{\nu'}{}^2E_1\left(\frac{\tau}{n_{\nu'}}\right)d\nu'$$
$$+\varpi\int_{\nu_1}^{\nu_2}\alpha_{\nu'}{}^2\Lambda_\tau\left\{\int_{\nu_1}^{\nu_2}\alpha_{\nu''}{}^2\,d\nu''\int_1^\infty J(t,u_{\nu''})\frac{du}{u}\;;\;n_{\nu'}\right\}d\nu',\quad(4.5)$$

where $x_{\nu'}=x/n_{\nu'}$, $u_{\nu''}=u/n_{\nu''}$. On subtracting $\frac{1}{2}\varpi J(0,\sigma_{\nu_0})$ times (4.5) from (4.4), we get a homogeneous Milne equation from which it follows that

$$\frac{\partial}{\partial\tau}J(\tau,\sigma_{\nu_0})+\sigma_{\nu_0}J(\tau,\sigma_{\nu_0})=\tfrac{1}{2}\varpi J(0,\sigma_{\nu_0})\int_{\nu_1}^{\nu_2}\alpha_{\nu'}{}^2\,d\nu'\int_1^\infty J(\tau,x_{\nu'})\frac{dx}{x}.\quad(4.6)$$

Let

$$R(s_\nu,\sigma_{\nu_0})=\int_0^\infty J(\tau,\sigma_{\nu_0})\,e^{-\tau s_\nu}\,d\tau,\quad(4.7)$$

where $s_\nu=s/n_\nu$. Multiply (4.6) by $e^{-\tau s_\nu}$, integrate with respect to τ over $(0,\infty)$ and then integrate the first term by parts. This gives us

$$-J(0,\sigma_{\nu_0})+s_\nu R(s_\nu,\sigma_{\nu_0})+\sigma_{\nu_0}R(s_\nu,\sigma_{\nu_0})=\tfrac{1}{2}\varpi J(0,\sigma_{\nu_0})\int_{\nu_1}^{\nu_2}\alpha_{\nu'}{}^2\,d\nu'\int_1^\infty R(s_\nu,x_{\nu'})\frac{dx}{x},$$

and therefore

$$(s_\nu+\sigma_{\nu_0})R(s_\nu,\sigma_{\nu_0})=J(0,\sigma_{\nu_0})\left\{1+\tfrac{1}{2}\varpi\int_{\nu_1}^{\nu_2}\alpha_{\nu'}{}^2\,d\nu'\int_1^\infty R(s_\nu,x_{\nu'})\frac{dx}{x}\right\}.\quad(4.8)$$

We next show that

$$R(s_\nu,\sigma_{\nu_0})=R(\sigma_{\nu_0},s_\nu).\quad(4.9)$$

The following method is convincing but is difficult to justify, because the integral on the left-hand sides of (4.10) and (4.11) (below) may not converge. A better method (due to Hopf) is used in (**4**) but it would make the analysis much longer.

Multiply (4.3) by $J(\tau,s_\nu)$ and integrate with respect to τ over $(0,\infty)$. Then

$$\int_0^\infty J(\tau,s_\nu)J(\tau,\sigma_{\nu_0})\,d\tau=\varpi\int_{\nu_1}^{\nu_2}\alpha_{\nu'}{}^2\,d\nu'\int_0^\infty J(\tau,s_\nu)\Lambda_\tau\{J(t,\sigma_{\nu_0})\,;\,n_{\nu'}\}\,d\tau+R(\sigma_{\nu_0},s_\nu).\quad(4.10)$$

Interchanging s and σ, ν and ν_0, we have

$$\int_0^\infty J(\tau,\sigma_{\nu_0})J(\tau,s_\nu)\,d\tau=\varpi\int_{\nu_1}^{\nu_2}\alpha_{\nu'}{}^2\,d\nu'\int_0^\infty J(\tau,\sigma_{\nu_0})\Lambda_\tau\{J(t,s_\nu)\,;\,n_{\nu'}\}\,d\tau+R(s_\nu,\sigma_{\nu_0}).\quad(4.11)$$

But the operator Λ is symmetrical in t and τ, and so

$$\int_0^\infty J(\tau,\sigma_{\nu_0})\Lambda_\tau\{J(t,s_\nu)\,;\,n_{\nu'}\}\,d\tau=\int_0^\infty J(\tau,s_\nu)\Lambda_\tau\{J(t,\sigma_{\nu_0})\,;\,n_{\nu'}\}\,d\tau.$$

* We shall usually invert orders of integration without commenting on it.

Hence (4.9) follows.

Next put $\tau = 0$ in (4.3). Then we get

$$\begin{aligned} J(0, \sigma_{\nu_0}) &= \tfrac{1}{2}\varpi \int_{\nu_1}^{\nu_2} \alpha_{\nu'}{}^2 \, d\nu' \int_0^\infty J(t, \sigma_{\nu_0}) E_1\left(\frac{t}{n_{\nu'}}\right) dt + 1 \\ &= 1 + \tfrac{1}{2}\varpi \int_{\nu_1}^{\nu_2} \alpha_{\nu'}{}^2 \, d\nu' \int_0^\infty J(t, \sigma_{\nu_0}) \, dt \int_1^\infty e^{-tx/n_{\nu'}} \frac{dx}{x} \\ &= 1 + \tfrac{1}{2}\varpi \int_{\nu_1}^{\nu_2} \alpha_{\nu'}{}^2 \, d\nu' \int_1^\infty R(x_{\nu'}, \sigma_{\nu_0}) \frac{dx}{x}. \end{aligned} \tag{4.12}$$

Using (4.9) and replacing σ_{ν_0} by s_ν, we have

$$J(0, s_\nu) = 1 + \tfrac{1}{2}\varpi \int_{\nu_1}^{\nu_2} \alpha_{\nu'}{}^2 \, d\nu' \int_1^\infty R(s_\nu, x_{\nu'}) \frac{dx}{x}. \tag{4.13}$$

By (4.13), (4.8) can be written

$$(s_\nu + \sigma_{\nu_0}) R(s_\nu, \sigma_{\nu_0}) = J(0, s_\nu) J(0, \sigma_{\nu_0}). \tag{4.14}$$

Then (4.13) becomes

$$J(0, s_\nu) = 1 + \tfrac{1}{2}\varpi J(0, s_\nu) \int_{\nu_1}^{\nu_2} \alpha_{\nu'}{}^2 \, d\nu' \int_1^\infty \frac{J(0, x_{\nu'})}{s_\nu + x_{\nu'}} \cdot \frac{dx}{x}. \tag{4.15}$$

Finally put $s = 1/\mu$, $x = 1/\mu'$, so that $s_\nu = 1/n_\nu \mu$, $x_{\nu'} = 1/n_{\nu'}\mu'$, and write

$$J(0, z^{-1}) = H(z) \tag{4.16}$$

and

$$H_\nu(\mu) = H(n_\nu \mu) = J(0, 1/n_\nu \mu). \tag{4.17}$$

Then (4.15) becomes

$$H_\nu(\mu) = 1 + \tfrac{1}{2}\varpi n_\nu \mu H_\nu(\mu) \int_{\nu_1}^{\nu_2} \alpha_{\nu'}{}^2 n_{\nu'} \, d\nu' \int_0^1 \frac{H_{\nu'}(\mu')}{n_\nu \mu + n_{\nu'}\mu'} \, d\mu',$$

and by (2.6), this can be written

$$H_\nu(\mu) = 1 + \tfrac{1}{2} n_\nu \mu H_\nu(\mu) \int_{\nu_1}^{\nu_2} \alpha_{\nu'}(1 - \lambda_{\nu'}) \, d\nu' \int_0^1 \frac{H_{\nu'}(\mu')}{n_\nu \mu + n_{\nu'}\mu'} \, d\mu'. \tag{4.18}$$

Also from (4.14) and (4.17), we have

$$R\left(\frac{1}{n_\nu \mu}, \frac{1}{n_{\nu_0}\mu'}\right) = \frac{n_\nu n_{\nu_0} \mu\mu' H_\nu(\mu) H_{\nu_0}(\mu')}{n_\nu \mu + n_{\nu_0}\mu'}. \tag{4.19}$$

If we put $n_\nu \mu = x$, $n_{\nu'}\mu' = x'$ in (4.18) and use (4.17), it becomes

$$H(x) = 1 + \tfrac{1}{2} x H(x) \int_{\nu_1}^{\nu_2} \frac{\alpha_{\nu'}(1 - \lambda_{\nu'})}{n_{\nu'}} \, d\nu' \int_0^{n_{\nu'}} \frac{H(x')}{x + x'} \, dx'. \tag{4.20}$$

But from (2.2) and (2.5)

$$\alpha_{\nu'}(1 - \lambda_{\nu'})/n_{\nu'} = (1 - \epsilon)\eta_{\nu'}{}^2 \Big/ \int_{\nu_1}^{\nu_2} \eta_{\nu'} \, d\nu' = (1 - \epsilon)\sigma_{\nu'}{}^2 \Big/ \kappa \int_{\nu_1}^{\nu_2} \sigma_{\nu'} \, d\nu'$$

and hence, if we take $\int_{\nu_1}^{\nu_2} \sigma_{\nu'} \, d\nu' = 1$ (as was done in (3)) and $\epsilon = 0$, (4.20) becomes

$$H(x) = 1 + \frac{1}{2\kappa} x H(x) \int_{\nu_1}^{\nu_2} \sigma_{\nu'}{}^2 \, d\nu' \int_0^{n_{\nu'}} \frac{H(x')}{x + x'} \, dx', \tag{4.21}$$

which (with $\nu_1 = 0$, $\nu_2 = \infty$) is equation (29) of (3). This verifies the correctness of the H-equation found in (3) and also in (8) and (9). We shall use the H-equation in the form (4.18).

From (4.18) it is easy to deduce that

$$\tfrac{1}{2} \int_{\nu_1}^{\nu_2} \alpha_{\nu'}(1 - \lambda_{\nu'}) \, d\nu' \int_0^1 H_{\nu'}(\mu') \, d\mu' = 1 - \left\{ \int_{\nu_1}^{\nu_2} \alpha_{\nu'} \lambda_{\nu'} \, d\nu' \right\}^{1/2}. \tag{4.22}$$

This is proved by the method of (6), Chap. V, Theorem I, to which it is equivalent.

From (4.18) and (4.22) we have the following alternative form for the H-equation:

$$\frac{1}{H_\nu(\mu)} = \left\{\int_{\nu_1}^{\nu_0} \alpha_{\nu'}\lambda_{\nu'}\,d\nu'\right\}^{1/2} + \tfrac{1}{2}\int_{\nu_1}^{\nu_0} \alpha_{\nu'}(1-\lambda_{\nu'})\,d\nu' \int_0^1 \frac{n_{\nu'}\mu' H_{\nu'}(\mu')}{n_\nu\mu + n_{\nu'}\mu'}\,d\mu'. \quad (4.23)$$

We shall need (4.22) and (4.23) later.

5. *The law of diffuse reflection.*—Having obtained the H-equation, we now have to show that a scattering function $S(\nu, \nu', \mu, \mu')$ exists which is given by

$$S(\nu, \nu', \mu, \mu') = \frac{(1-\lambda_\nu)\alpha_{\nu'} n_{\nu'}\mu\mu' H_\nu(\mu) H_{\nu'}(\mu')}{n_\nu\mu + n_{\nu'}\mu'}, \quad (5.1)$$

or by

$$S(\nu, \nu', \mu, \mu') = \frac{(1-\lambda_\nu)\alpha_{\nu'}}{n_\nu} R\left(\frac{1}{n_\nu\mu}, \frac{1}{n_{\nu'}\mu'}\right), \quad (5.2)$$

and which is such that the incident intensity $f_\nu(\mu')$ is related to the emergent intensity $I_\nu^*(0, \mu)$ by the equation

$$I_\nu^*(0, \mu) = \frac{1}{2\mu}\int_{\nu_1}^{\nu_0} d\nu' \int_0^1 f_{\nu'}(\mu') S(\nu, \nu', \mu, \mu')\,d\mu'. \quad (5.3)$$

Equation (5.1) is equivalent to equation (33) of (3). By (2.6) and (4.9), $S(\nu, \nu', \mu, \mu')$ satisfies the relation

$$S(\nu, \nu', \mu, \mu') = S(\nu', \nu, \mu', \mu) \quad (5.4)$$

as it should.

We now turn to equation (3.5). Since $\mathfrak{J}_\nu(\tau)/(1-\lambda_\nu)$ is independent of ν, we can write

$$\mathfrak{J}_\nu(\tau) = (1-\lambda_\nu)J(\tau). \quad (5.5)$$

Then, by (2.6),

$$J(\tau) = \varpi\int_{\nu_1}^{\nu_0} \alpha_{\nu'}\Lambda_{\nu'}\{J(t);\, n_{\nu'}\}\,d\nu' + \tfrac{1}{2}\int_{\nu_1}^{\nu_0} \alpha_{\nu'}\,d\nu' \int_0^1 f_{\nu'}(\mu')\,e^{-\tau/n_{\nu'}\mu'}\,d\mu'.$$

In the last integral put $\mu' = 1/x$. Then

$$J(\tau) = \varpi\int_{\nu_1}^{\nu_0} \alpha_{\nu'}\Lambda_{\nu'}\{J(t);\, n_{\nu'}\}\,d\nu' + \tfrac{1}{2}\int_{\nu_1}^{\nu_0} \alpha_{\nu'}\,d\nu' \int_1^\infty f_{\nu'}(x^{-1})\,e^{-\tau x_{\nu'}}\frac{dx}{x^2}, \quad (5.6)$$

where $x_{\nu'} = x/n_{\nu'}$. Differentiating, we have by (3.7)

$$J'(\tau) = \varpi\int_{\nu_1}^{\nu_0} \alpha_{\nu'}\Lambda_{\nu'}\{J'(t);\, n_{\nu'}\}\,d\nu' + \tfrac{1}{2}\varpi J(0)\int_{\nu_1}^{\nu_0} \alpha_{\nu'}E_1\left(\frac{\tau}{n_{\nu'}}\right)d\nu'$$
$$- \tfrac{1}{2}\int_{\nu_1}^{\nu_0} \frac{\alpha_{\nu'}}{n_{\nu'}}\,d\nu' \int_1^\infty f_{\nu'}(x^{-1})\,e^{-\tau x_{\nu'}}\frac{dx}{x}. \quad (5.7)$$

The second integral on the right-hand side is given by (4.5). To get the third integral, multiply the auxiliary equation (4.3) by $f_{\nu_0}(\sigma^{-1})/\sigma$ and integrate over $(1, \infty)$. This gives us

$$\int_1^\infty J(\tau, \sigma_{\nu_0}) f_{\nu_0}(\sigma^{-1})\frac{d\sigma}{\sigma} = \varpi\int^{\nu_0} \alpha_{\nu'}\Lambda_{\nu'}\left\{\int_1^\infty J(t, \sigma_{\nu_0}) f_{\nu_0}(\sigma^{-1})\frac{d\sigma}{\sigma};\; n_{\nu'}\right\} d\nu'$$
$$+ \int_1^\infty J_{\nu_0}(\sigma^{-1})\,e^{-\tau\ \nu_0}\frac{d\sigma}{\sigma},$$

and hence

$$\tfrac{1}{2}\int_{\nu_1}^{\nu_0}\frac{\alpha_{\nu_0}}{n_{\nu_0}}d\nu_0\int_1^\infty J(\tau,\sigma_{\nu_0})f_{\nu_0}(\sigma^{-1})\frac{d\sigma}{\sigma}$$
$$=\tfrac{1}{2}\int_{\nu_1}^{\nu_0}\frac{\alpha_{\nu_0}}{n_{\nu_0}}d\nu_0\int_1^\infty f_{\nu_0}(\sigma^{-1})\,e^{-\tau\sigma_{\nu_0}}\frac{d\sigma}{\sigma}+\tfrac{1}{2}\varpi\int_{\nu_1}^{\nu_0}\frac{\alpha_{\nu_0}}{n_{\nu_0}}d\nu_0\int_{\nu_1}^{\nu_0}\alpha_{\nu'}{}^2$$
$$\times\Lambda_\tau\left\{\int_1^\infty J(t,\sigma_{\nu_0})f_{\nu_0}(\sigma^{-1})\frac{d\sigma}{\sigma}\;;\;n_{\nu'}\right\}d\nu'.$$

Inverting the orders of integration in the last term and changing the variables σ and ν_0, we have

$$\tfrac{1}{2}\int_{\nu_1}^{\nu_0}\frac{\alpha_{\nu'}}{n_{\nu'}}d\nu'\int_1^\infty J(\tau,x_{\nu'})f_{\nu'}(x^{-1})\frac{dx}{x}$$
$$=\tfrac{1}{2}\int_{\nu_1}^{\nu_0}\frac{\alpha_{\nu'}}{n_{\nu'}}d\nu'\int_1^\infty f_{\nu'}(x^{-1})\,e^{-\tau x_{\nu'}}\frac{dx}{x}+\tfrac{1}{2}\varpi\int_{\nu_1}^{\nu_0}\alpha_{\nu'}{}^2$$
$$\times\Lambda_\tau\left\{\int_{\nu_1}^{\nu_0}\frac{\alpha_{\nu''}}{n_{\nu''}}d\nu''\int_1^\infty J(t,u_{\nu''})f_{\nu''}(u^{-1})\frac{du}{u}\;;\;n_{\nu'}\right\}d\nu'. \quad (5.8)$$

On adding this to (5.7) and subtracting $\frac{1}{2}\varpi J(0)$ times (4.5), we obtain a homogeneous Milne equation and we therefore have

$$J'(\tau)=\tfrac{1}{2}\varpi J(0)\int_{\nu_1}^{\nu_0}\alpha_{\nu'}{}^2d\nu'\int_1^\infty J(\tau,x_{\nu'})\frac{dx}{x}-\tfrac{1}{2}\int_{\nu_1}^{\nu_0}\frac{\alpha_{\nu'}}{n_{\nu'}}d\nu'\int_1^\infty J(\tau,x_{\nu'})f_{\nu'}(x^{-1})\frac{dx}{x}. \quad (5.9)$$

Let

$$j(s)=\mathfrak{L}_s\{J(\tau)\}=s\int_0^\infty J(\tau)\,e^{-s\tau}\,d\tau. \quad (5.10)$$

Operate on (5.9) by $s_\nu{}^{-1}\mathfrak{L}_{\tau_\nu}$ and then integrate by parts on the left. This gives us, by (4.7),

$$-J(0)+j(s_\nu)=\tfrac{1}{2}\varpi J(0)\int_{\nu_1}^{\nu_0}\alpha_{\nu'}{}^2d\nu'\int_1^\infty R(s_\nu,x_{\nu'})\frac{dx}{x}$$
$$-\tfrac{1}{2}\int_{\nu_1}^{\nu_0}\frac{\alpha_{\nu'}}{n_{\nu'}}d\nu'\int_1^\infty R(s_\nu,x_{\nu'})f_{\nu'}(x^{-1})\frac{dx}{x},$$

and by (4.13), this can be written

$$j(s_\nu)=J(0)J(0,s_\nu)-\tfrac{1}{2}\int_{\nu_1}^{\nu_0}\frac{\alpha_{\nu'}}{n_{\nu'}}d\nu'\int_1^\infty R(s_\nu,x_{\nu'})f_{\nu'}(x^{-1})\frac{dx}{x}. \quad (5.11)$$

We now require $J(0)$. A great deal of labour can be saved by the following physical argument. Since the only solution of (3.8) under our conditions is $\phi_\nu(\tau)\equiv 0$, the source function $\mathfrak{J}_\nu(\tau)$ given by (3.5) is entirely due to the incident radiation of intensity $f_\nu(\mu')$. This will have no effect on $\mathfrak{J}_\nu(\tau)$ for large τ (the corresponding term on the right-hand side of (3.5) tends to zero as $\tau\to\infty$) and so $\mathfrak{J}_\nu(\tau)\to 0$ as $\tau\to\infty$. Since, by (5.5) and (5.10), $(1-\lambda_\nu)j(s_\nu)$ is the Laplace transform of $\mathfrak{J}_\nu(\tau)$, it follows from a well-known theorem that

$$j(s_\nu)\to 0 \quad \text{as} \quad s\to 0. \quad (5.12)$$

From (4.14) we have

$$\lim_{s\to 0}R(s_\nu,x_{\nu'})=J(0,0)J(0,x_{\nu'})/x_{\nu'}$$

and hence (5.11) gives, on letting $s\to 0$,

$$J(0)=\tfrac{1}{2}\int_{\nu_1}^{\nu_0}\alpha_{\nu'}d\nu'\int_1^\infty J(0,x\nu')f_{\nu'}(x^{-1})\frac{dx}{x^2}.$$

On putting $x=1/\mu'$ and using (4.17), this becomes

$$J(0)=\tfrac{1}{2}\int_{\nu_1}^{\nu_2}\alpha_\nu\,d\nu'\int_0^1 H_\nu(\mu')f_\nu(\mu')\,d\mu'. \tag{5.13}$$

To establish this equation mathematically, we put $\tau=0$ in (5.6) and then proceed as in (4.12). This gives us an equation involving

$$\int_{\nu_1}^{\nu_2}\alpha_\nu{}^3 n_\nu\,d\nu'\int_1^\infty j(x_\nu)\frac{dx}{x^3}.$$

On substituting the value of $j(x_\nu)$ given by (5.11), we get an equation giving $J(0)$ in terms of the functions $J(0,s_\nu)$ and $f_\nu(x^{-1})$. On converting the former into $H_\nu(\mu)$ and using (4.22) and (4.23) in the subsequent reduction, we eventually arrive at (5.13).

Put $s=1/\mu$, $x=1/\mu'$ in (5.11). Then, by (4.19),

$$j\left(\frac{1}{n_\nu\mu}\right)=J(0)H_\nu(\mu)-\tfrac{1}{2}n_\nu\mu H_\nu(\mu)\int_{\nu_1}^{\nu_2}\alpha_{\nu'}\,d\nu'\int_0^1\frac{H_{\nu'}(\mu')f_{\nu'}(\mu')}{n_\nu\mu+n_{\nu'}\mu'}\,d\mu'$$

and hence by (5.13)

$$\begin{aligned}j\left(\frac{1}{n_\nu\mu}\right)&=\tfrac{1}{2}H_\nu(\mu)\int_{\nu_1}^{\nu_2}\alpha_{\nu'}\,d\nu'\int_0^1\frac{n_{\nu'}\mu' H_{\nu'}(\mu')f_{\nu'}(\mu')}{n_\nu\mu+n_{\nu'}\mu'}\,d\mu'\\&=\frac{1}{2\mu}\int_{\nu_1}^{\nu_2}\frac{\alpha_{\nu'}}{n_{\nu'}}\,d\nu'\int_0^1 R\left(\frac{1}{n_\nu\mu},\frac{1}{n_{\nu'}\mu'}\right)f_{\nu'}(\mu')\,d\mu'.\end{aligned} \tag{5.14}$$

But by (3.3) and (5.5)

$$I_\nu{}^*(0,\mu)=(1-\lambda_\nu)j\left(\frac{1}{n_\nu\mu}\right), \tag{5.15}$$

and so (5.14) becomes

$$I_\nu{}^*(0,\mu)=\frac{1}{2\mu}\int_{\nu_1}^{\nu_2}d\nu'\int_0^1 S(\nu,\nu',\mu,\mu')f_{\nu'}(\mu')\,d\mu',$$

where $S(\nu,\nu',\mu,\mu')$ is given by (5.2). This establishes the law of diffuse reflection (5.3).

6. *Concluding note.*—In (3) and (5), all work subsequent to the establishment of the scattering function is based on the H-equation and the law of diffuse reflection. Since these have been re-established, the consequent investigations are unchanged.

The forms into which formulae have been put in (5) are better than those used for the corresponding formulae in (3). Since the notation used in this paper is the "continuous" analogue of that used in (5), the better forms for the non-coherent case can be written down at sight by replacing sums over $\nu_1, \ldots, \nu_k$ by integrals over $\nu_1 \leqslant \nu \leqslant \nu_2$.

St Hugh's College,
Oxford:
1955 *May* 25.

Note added on proof.—I am indebted to Dr Stibbs for the correction of an error in Section 1. He has also drawn my attention to a paper by S. Ueno (*Contributions from the Inst. of Astroph., Univ. of Kyoto*, No. 58, 1955) in which ideas somewhat similar to those used here can be traced. Reference is made in this paper to "the extension of Ambartsumian's first method developed in our subsequent paper". This subsequent paper has not yet appeared, but it is possible that Ueno and I have developed Ambartsumian's method in the same way.

References

(1) V. A. Ambartsumian, *C.R. (Doklady) Acad. Sci. U.R.S.S.*, **38**, 257, 1943.
(2) V. A. Ambartsumian, *Ast. J. of the Soviet Union*, **19**, 30, 1942.
(3) I. W. Busbridge, *M.N.*, **113**, 52, 1953.
(4) I. W. Busbridge, *Quart. J.* (Oxford) (2), **6**, 218, 1955.
(5) I. W. Busbridge and D. W. N. Stibbs, *M.N.*, **114**, 2, 1954.
(6) S. Chandrasekhar, *Radiative Transfer*, Oxford, 1950.
(7) V. Kourganoff, *Basic Methods in Transfer Problems*, Oxford, 1952.
(8) V. V. Sobolev, *R.A.J.*, **26**, 129, 1949.
(9) V. V. Sobolev, *R.A.J.*, **31**, 231, 1954.
(10) D. W. N. Stibbs, *M.N.*, **113**, 493, 1953.

HYDROMAGNETIC WAVES IN IONIZED GAS

J. H. Piddington

(Received 1955 May 3)

Summary

There is strong evidence that both magnetic fields and ionized gas are widespread in our galaxy with powerful concentrations around stars and, in particular, the Sun. In such regions a localized disturbance will inevitably cause one or more of the different types of hydromagnetic waves. These are combined *mechanical* and *electromagnetic* waves, and the forces involved may far outweigh gravitational and other forces present. The electromagnetic forces, hitherto largely neglected, should be considered in many astrophysical problems.

Most investigators of HM waves in gases have, for simplicity, assumed either perfect or isotropic conductivity which is seldom justified and completely hides the nature of the *electromagnetic* wave. In the present paper a general theory of weak HM waves in anisotropically conducting gas is developed. There are three such HM waves and for each the velocity, absorption, magnetic and electric fields, the Poynting vector and the gas velocity are found.

The physical nature of some HM waves is discussed; these include, as limiting cases, radio waves in ionized gas and *in vacuo*. The effect of Maxwell's displacement current on HM waves is found to be generally small, but it does cause a space-charge electric wave to accompany the HM wave and this could have important consequences.

1. *Introduction.*—Hydromagnetic (or magneto-hydrodynamic) waves require, like all other waves, a medium with inertia and a restoring force opposing displacement from the steady state. In ionized gas, permeated by a steady magnetic field, the inertia is provided by the gas as a whole, and the restoring force by a combination of electromagnetic and gas pressure forces. This medium is triply refracting allowing three distinct hydromagnetic (HM) waves to propagate at any given frequency. Two of these degenerate to radio waves when the frequency becomes so high that heavy ion motion is negligible; the inertia is then provided by the electrons and by the electromagnetic field itself. The third wave degenerates to a sound wave when the magnetic field is sufficiently weak. The restoring force is then provided by the gas pressure only.

Recent evidence suggests that the galaxy is permeated (and perhaps surrounded) by a magnetic field having powerful concentrations in and near stars and the Sun, and perhaps for considerable distances around old supernovae. Ionized gas is also widespread and when present with a magnetic field conditions are such that any localized disturbance *must* cause HM waves. The electromagnetic forces associated with such HM disturbances may be large; in many regions they may far outweigh gravitational forces and so must be introduced into a wide range of astrophysical problems. For example, their effects in parts of the solar atmosphere are of major importance (**1**). They also constitute the most likely source of cosmic rays and, in turn, of galactic radio noise.

The general theory of HM waves is mathematically difficult and advances have been made by assuming simplifying conditions. Alfvén (**2, 3**) dealt with waves in a homogeneous incompressible fluid, and Walén (**4**) and others extended the work to inhomogeneous fluids. HM waves in gases with infinite (**5, 6, 7**) and isotropic (**8**) electrical conductivity have been discussed. However, in most regions of astrophysical interest the gas has neither infinite nor isotropic conductivity and the adoption of these simplifying assumptions completely hides the nature of the electromagnetic field associated with the HM wave. Aström (**9**) and Ginzburg (**10**) have further extended the theory but the former neglects collisions between gas particles (and hence gas pressure effects), while the latter only considers waves propagated along the steady magnetic field. A reasonably full HM theory is still lacking.

In a moving, isotropically conducting medium, Maxwell's equations of the electromagnetic field may be combined with the conductivity equation (Ohm's law) to yield the well-known formula:

$$\nabla^2 \boldsymbol{H} = 4\pi\sigma\left\{\frac{\partial \boldsymbol{H}}{\partial t} - \operatorname{curl}(\boldsymbol{v}\times\boldsymbol{H})\right\} + \frac{1}{c^2}\frac{\partial^2 \boldsymbol{H}}{\partial t^2} \qquad (1)$$

where $\boldsymbol{H}$ is the magnetic field, σ and $\boldsymbol{v}$ the conductivity and velocity of the medium and c the velocity of light, electromagnetic units being used. When $\boldsymbol{v}=0$ this reduces to the "telegraph equation" of radio waves. When $\boldsymbol{v}$ depends on electromagnetic forces it leads easily to a description of HM waves.

When the medium is anisotropically conducting the conductivity is no longer a scalar quantity as above, but a tensor. An equation analogous to (1) has been derived for such a medium (**11**) and this may be used to investigate any weak electromagnetic disturbance in ionized gas. In the present paper it is used on plane hydromagnetic waves. At the low frequencies concerned displacement current may usually be neglected.

2. *The electromagnetic and hydromagnetic equations.*—The equations of the magnetic field associated with a weak disturbance in any medium have been given (ref. **11**, equations (14), (16)). If the variable parts of $\boldsymbol{H}$ can be determined from those equations, then the remainder of the electromagnetic field follows from Maxwell's equations. Plane waves are assumed, and without loss of generality the wave normal is placed in the xz plane and the steady magnetic field H_0 is made parallel to the Oz axis of a right-handed system. Since we are restricted to weak waves, the product of any two variables is always small and may be neglected. The equations may then be written:

$$\frac{\partial^2 H_x}{\partial x^2} + \frac{\partial^2 H_x}{\partial z^2} = 4\pi\sigma_3\left(\frac{\partial H_x}{\partial t} - H_0\frac{\partial v_x}{\partial z}\right) - \frac{\sigma_2}{\sigma_1}\frac{\partial^2 H_y}{\partial z^2} \qquad (2)$$

$$\frac{\sigma_3}{\sigma_0}\frac{\partial^2 H_y}{\partial x^2} + \frac{\partial^2 H_y}{\partial z^2} = 4\pi\sigma_3\left(\frac{\partial H_y}{\partial t} - H_0\frac{\partial v_y}{\partial z}\right) + \frac{\sigma_2}{\sigma_1}\left(\frac{\partial^2 H_x}{\partial z^2} - \frac{\partial^2 H_z}{\partial x \partial z}\right) \qquad (3)$$

$$\frac{\partial^2 H_z}{\partial x^2} + \frac{\partial^2 H_z}{\partial z^2} = 4\pi\sigma_3\left(\frac{\partial H_z}{\partial t} + H_0\frac{\partial v_x}{\partial x}\right) + \frac{\sigma_2}{\sigma_1}\frac{\partial^2 H_y}{\partial x \partial z}, \qquad (4)$$

where σ_0, σ_1 and σ_2 are the three components of the conductivity tensor, σ_0 being the component along $\boldsymbol{H}_0$, σ_1 perpendicular to $\boldsymbol{H}_0$ and in the plane of the electric field (the "direct" conductivity) and σ_2 perpendicular to $\boldsymbol{H}_0$ and to the electric field (the Hall conductivity). The composite conductivity σ_3 is given by $\sigma_3 = \sigma_1 + \sigma_2^2/\sigma_1$.

These equations apply for any kind of medium, including gas in which the heavy ions as well as the electrons contribute appreciably to the conductivity (as in ionospheric region E). However, in most applications of astrophysical interest we are concerned with gas in which only electron conductivity is important. In this case the value of σ_3/σ_0 is unity and equation (3) is correspondingly simpler. We restrict ourselves to this type of gas for the present, but it should be noted that the results apply for any gas when the direction of propagation is along the magnetic field (so that $\partial/\partial x = 0$). With this restriction it is seen that equations (2) to (4) are the components of equation (1) with σ_3 replacing σ and the addition of a term $(\sigma_2/\sigma_1)(\partial/\partial z)\operatorname{curl} \boldsymbol{H}$ to the right-hand side. This last term changes the magnetic polarization of the waves from plane to elliptically polarized except in the special case of propagation across the magnetic field ($\partial/\partial_z = 0$), when the term vanishes.

The hydrodynamic equations required to determine $\boldsymbol{v}$ are those of momentum, continuity and the "gas equation" relating pressure density and temperature. The momentum equation contains an electromagnetic term $\boldsymbol{j} \times \boldsymbol{H}$ or

$$\left(\frac{1}{4\pi} \operatorname{curl} \boldsymbol{H}\right) \times \boldsymbol{H}$$

(see, for example, ref. **11**, section 6). For weak disturbances the hydrodynamic equations may be combined to give

$$\frac{\partial^2 \boldsymbol{v}}{\partial t^2} - U^2 \operatorname{grad} \operatorname{div} \boldsymbol{v} + \frac{V^2}{H_0^2} \frac{\partial}{\partial t} (\boldsymbol{H}_0 \times \operatorname{curl} \boldsymbol{H}) = 0, \tag{5}$$

where $U^2 = \partial\theta/\partial\rho$ is the square of the velocity of sound in the gas, θ and ρ being the pressure and mass density and $V^2 = H_0^2/4\pi\rho$.

Equations (2), (3) and (4) may now be combined with (5) to determine the values of $\boldsymbol{H}$ and $\boldsymbol{v}$. Let us assume wave solutions of the form $\exp i(pt - k\xi)$ propagated along an axis $O\xi$ at an angle ψ to Oz. Let H_x, H_y and H_z now represent the maximum values of the variable components of $\boldsymbol{H}$. An examination of the equations shows that in all cases $H_z = -H_x \tan\psi$ so that the perturbation magnetic field always lies in the phase planes and H_z may be eliminated from the field equations.

Substituting $\partial/\partial x = -ik\sin\psi$, $\partial/\partial z = -ik\cos\psi$ and $\partial/\partial t = ip$ we have the set of equations required to investigate hydromagnetic waves. These are simplified by writing ϵ_0 for $4\pi\sigma_0/p$ (or $4\pi\sigma_3/p$):

$$(k^2 + i\epsilon_0 p^2)H_x + i\epsilon_0 H_0 \cos\psi k p v_x + \sigma_2/\sigma_1 \cos^2\psi k^2 H_y = 0 \tag{6}$$

$$(k^2 + i\epsilon_0 p^2)H_y + i\epsilon_0 H_0 \cos\psi k p v_y - \sigma_2/\sigma_1 k^2 H_x = 0 \tag{7}$$

$$(p^2 - U^2 \sin^2\psi k^2)v_x - U^2 \sin\psi \cos\psi k^2 v_z + \frac{V^2 k p}{H_0 \cos\psi} H_x = 0 \tag{8}$$

$$p v_y + \frac{V^2 \cos\psi k}{H_0} H_y = 0 \tag{9}$$

$$(p^2 - U^2 \cos^2\psi k^2)v_z - U^2 \sin\psi \cos\psi k^2 v_x = 0. \tag{10}$$

Eliminating v_x, v_y and v_z we have:

$$\left\{k^2 + i\epsilon_0 p^2 - i\epsilon_0 V^2 k^2 \left(\frac{p^2 - U^2 \cos^2\psi k^2}{p^2 - U^2 k^2}\right)\right\} H_x + \sigma_2/\sigma_1 \cos^2\psi k^2 H_y = 0 \tag{11}$$

and

$$(k^2 + i\epsilon_0 p^2 - i\epsilon_0 V^2 \cos^2\psi k^2)H_y - \sigma_2/\sigma_1 k^2 H_x = 0. \tag{12}$$

These two equations may be used to determine the velocity, absorption and polarization of any hydromagnetic wave, subject to the restrictions already mentioned. When H_x and H_y are eliminated, the dispersion equation is obtained. This is of the third degree in k^2 showing that, in general, three pairs of waves may be present for any value of ψ; each pair comprises two similar waves moving in opposite directions. The medium is triply refracting, introducing an additional pair of waves and an additional degree of complication above those given by the magneto-ionic theory.

The dispersion equation cannot, in general, be factorized and its discussion is difficult except in some special circumstances. In the following sections the properties of waves are derived for the following restricted conditions:

$$(a)\ \psi = 0$$
$$(b)\ \psi = \pi/2$$
$$(c)\ V \gg U$$
$$(d)\ U \gg V.$$

3. *Waves propagated along and perpendicular to the magnetic field.*—A glance at equation (11) indicates that a great simplification occurs when propagation is along the steady magnetic field, that is when $\psi = 0$. The equation then factorizes, one factor being $p^2 - U^2k^2$ which corresponds to pure sound waves with velocities $\pm U$. These waves correspond to gas motion along the magnetic field, there being no interaction with this field and no electromagnetic wave*. No attenuation is indicated because friction within the gas has been neglected.

The two pairs of hydromagnetic waves may now be studied by the use of equation (12) and the simplified form of equation (11). This has been done in an earlier paper (ref. **11**): the waves along Oz are circularly polarized and have velocities $V \pm \sigma_2 p/8\pi\sigma_1\sigma_0 V$, in agreement with Cowling (**12**). It may be noted that these results apply for any type of gas including lightly ionized gas.

It is desirable to compare the dispersion equation for the case $\psi = 0$ with that found by Ginzburg who used a different treatment. In terms of transport parameters the equation is

$$\frac{k^2}{p^2} = \left(V^2 \pm \frac{\sigma_2 p}{4\pi\sigma_1\sigma_0} + \frac{ip}{4\pi\sigma_0} \right)^{-1}. \tag{13}$$

When the collision frequency is much less than the electron gyro-frequency, and when only electron conductivity need be considered, $\sigma_0 \simeq \sigma_2^2/\sigma_1$. Substituting ion constants for σ_1, σ_2, σ_0 and V (see, for example, **13**) equation (13) becomes

$$\frac{k^2c^2}{p^2} = \frac{-4\pi ne^2c^2}{mp(-\omega\pm\Omega p^{-1} \pm \omega - i\nu)}, \tag{14}$$

where n, e, m, ω and ν are the electron number density, charge, mass, angular gyro-frequency and collision frequency and Ω is the heavy-ion angular gyro-frequency. The equation is identical with that of Ginzburg except that a term unity is omitted on the right hand side. This term is due to the displacement current which we have assumed negligible. The corresponding equation for a lightly ionized gas may be found by substituting the appropriate values for the conductivity components.

* Strictly speaking, the sound wave will be accompanied by an electric space-charge or plasma wave. To study the electric field account must be taken of displacement current.

The case of transmission across the magnetic field has been considered previously (**11**). Two pairs of waves are disclosed for which H_x and v_x are the variables. In addition there is another pair, not disclosed by equations (11) and (12), for which the electric field and current lie along Oz and the magnetic variation along Oy. For this wave there is no mechanical reaction between the gas and magnetic field.

4. *Medium for which $V \gg U$. (a) The "magnetic" waves.*—When the "magnetic pressure" in the medium ($H_0^2/8\pi$) is much greater than the gas pressure the elastic properties due to the latter may be neglected; the restoring force opposing deformation of the homogeneous medium is almost entirely electromagnetic.

In equation (11) let $U=0$ for the time being. It may then be solved with equation (12) and the values of p/k corresponding to the two waves determined; the values are found to be of the order V. Now let U have a finite value provided $V \gg U$: the values of p/k will remain of order V so that $p/k \gg U$ and the equation takes the simplified form:

$$(k^2 + i\epsilon_0 p^2 - i\epsilon_0 V^2 k^2) H_x + (\sigma_2/\sigma_1) \cos^2\psi k^2 H_y = 0. \tag{15}$$

To find the polarization of the two pairs of hydromagnetic waves introduce $R = H_y/H_p$ where H_p is the perturbation magnetic field in the xz plane, so that

$$H_p = \frac{H_x}{\cos\psi} \quad \text{and} \quad \frac{H_y}{H_x} = \frac{R}{\cos\psi}.$$

Eliminating H_x and H_y from equations (12) and (15) we have

$$R^2 - \frac{i\epsilon_0\sigma_1 V^2 \sin^2\psi}{\sigma_2 \cos\psi} R + 1 = 0. \tag{16}$$

Since we are dealing with a gas in which only electron conductivity is important, $\sigma_0\sigma_1/\sigma_2 = ne/H_0$ (see **13**) and so

$$\frac{\epsilon_0\sigma_1 V^2}{\sigma_2} = \frac{\Omega}{p}. \tag{17}$$

The two roots of equation (16), corresponding to the polarization of the two waves concerned, are:

$$R = \frac{i}{2}\left\{\frac{\Omega \sin^2\psi}{p\cos\psi} \pm \left(\frac{\Omega^2 \sin^4\psi}{p^2\cos^2\psi} + 4\right)^{1/2}\right\}. \tag{18}$$

The waves are elliptically polarized, the major axis of the ellipse lying along Oy for the positive sign (called the ordinary ray) and in the xz plane for the negative sign (called the extraordinary ray). Values of the polarization function R/i are plotted in Fig. 1 for all angles of propagation and for values of Ω/p between* 10^{-1} and 10^4. In the wide range of examples when $\Omega/p \gtrsim 10^3$ the waves are nearly plane polarized except for very small angles between the magnetic field and wave normal. For these waves, the longitudinal case as discussed by Ginzburg is not important; it only describes waves whose direction of propagation lies within a very small solid angle. For values of Ω/p between about 1 and 100 the polarization varies more steadily from circular to plane as ψ varies from 0 to $\pi/2$.

* Since we are dealing with a gas in which only electron conductivity is important, transport equations may be used for values of $\Omega/p < 1$, provided that $\omega/p \gg 1$. This has been discussed previously (**13**).

Substitute the two values of $H_y/H_x (= R/\cos\psi)$ from equation (18) in equations (12) or (15) and the dispersion equation is found:

$$p^2/k^2 = V^2\left\{1 - \frac{\sin^2\psi}{2} \mp \left(\frac{\sin^4\psi}{4} + \frac{p^2}{\Omega^2}\cos^2\psi\right)^{1/2}\right\} + \frac{i}{\epsilon_0} = V^2S^2 + i/\epsilon_0. \quad (19)$$

The wave velocity is VS and the absorption coefficient

$$\kappa = \frac{p}{2\epsilon_0 V^2 S^2} = \frac{p^2}{8\pi\sigma_0 V^2 S^2}. \quad (20)$$

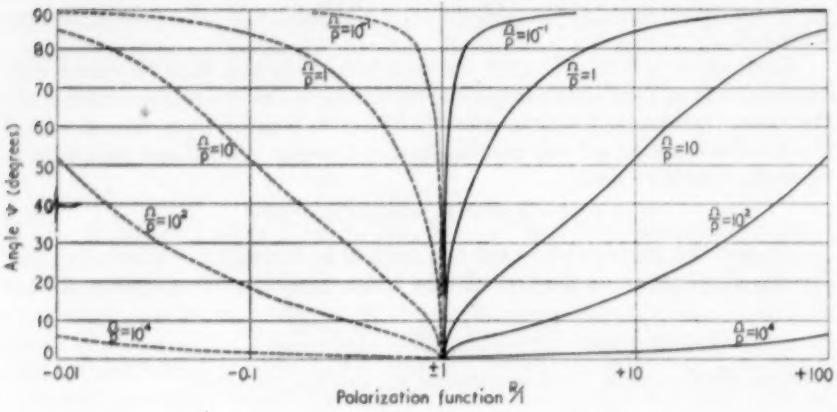


FIG. 1.—*The polarization $R/i = H_y/iH_p$ for values of Ω/p between 10^{-1} and 10^4, and for angles (ψ) between the steady magnetic field and the wave normal between 0° and 90°. Curves corresponding to positive values of Ri^{-1} (the ordinary ray) are full; curves corresponding to negative values of Ri^{-1} (the extraordinary ray) are dashed.*

The wave decays by a factor e^{-1} in a distance κ^{-1}. The velocity function:

$$S = \left\{1 - \frac{\sin^2\psi}{2} \pm \left(\frac{\sin^4\psi}{4} + \frac{p^2}{\Omega^2}\cos^2\psi\right)^{1/2}\right\}^{1/2} \quad (21)$$

is plotted in Fig. 2 for all angles of propagation and for values of Ω/p between 10^{-1} and infinity.

At low frequencies the term in p^2/Ω^2 may be neglected and we have

$$S = \cos\psi \quad \text{or} \quad 1.$$

Thus the velocity of the ordinary ray is $V\cos\psi$ and that of the extraordinary ray V. Although the results are for plane waves they should apply approximately for curved wave fronts, provided the radius of curvature is many wavelengths. In Fig. 3 the wave fronts of the waves are shown when the waves are caused by a disturbance in a confined region near the origin. The wave fronts are spheres* obtained by rotating the circles about the axis Oz.

At high frequencies the velocity of the ordinary ray approaches zero and when $\Omega/p = 1$ it is zero for all values of ψ. At higher frequencies the value of S is imaginary. The corresponding values of S for the extraordinary ray increase steadily above V.

* When displacement current is taken into account the ordinary ray wave front takes a rather different form in the region $\psi \to \pi/2$.

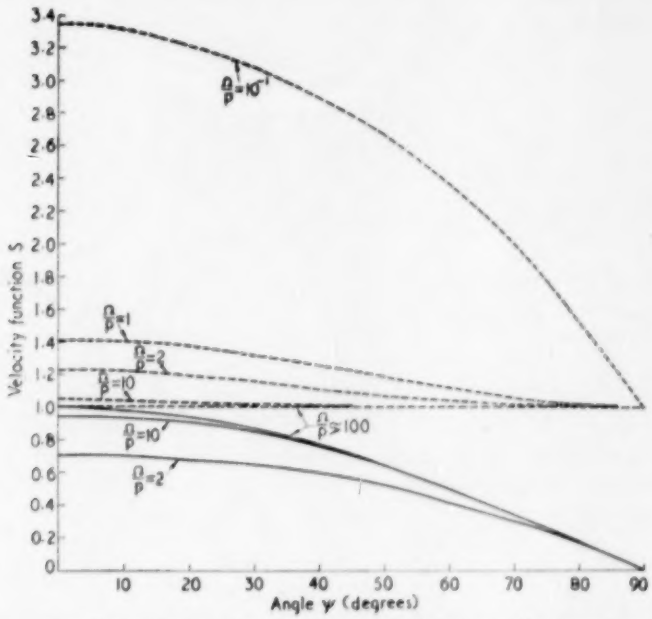


FIG. 2.—*The velocity function S (wave velocity is VS) for values of Ω/p between 10^{-1} and ∞ and for angles (ψ) between the steady magnetic field and the wave normal between 0° and 90°. The ordinary ray curves are full and the extraordinary ray curves are dashed. The former are not drawn for Ω/p ⩽ 1 since the corresponding values of S are everywhere zero or imaginary.*

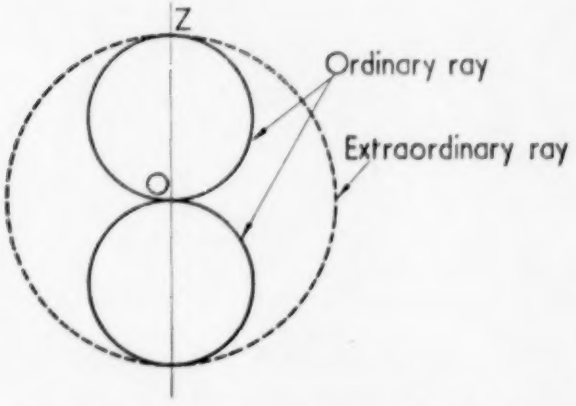


FIG. 3.—*Wave fronts of hydromagnetic waves originating in a confined region near O.*

The gas velocity components may be found from equations (8) to (10) and the dispersion equation. Neglecting the small effect of absorption we may write $p/k = VS$ so that

$$v_x = -\frac{V}{S\cos\psi}\cdot\frac{H_z}{H_0}$$

$$v_y = -\frac{V\cos\psi}{S}\cdot\frac{H_y}{H_0}$$

$$v_z = 0$$

$$\frac{v_y}{v_x} = \cos^2\psi\,\frac{H_y}{H_x} = R\cos\psi.$$

Thus a small element of the gas moves in an elliptical orbit under the influence of either the ordinary or extraordinary waves. For the former the major axis of the ellipse lies along Oy and for the latter along Ox. At low frequencies $(\Omega/p \gg 1)$

$$\frac{v_x}{v_y} = \mathrm{i}\,\frac{\Omega}{p}\sin^2\psi \quad \text{or} \quad \mathrm{i}\,\frac{p}{\Omega}\cot^2\psi.$$

Thus the movement is nearly along Oy or nearly along Ox for the two waves.

The electric polarization of the waves and the direction of current flow may be found from the equations

$$\operatorname{curl}\boldsymbol{E} = \frac{-\partial\boldsymbol{H}}{\partial t} \tag{22}$$

$$\operatorname{curl}\boldsymbol{H} = 4\pi\boldsymbol{j}. \tag{23}$$

From the x and z components of equation (23) we find

$$\frac{j_z}{j_x} = -\tan\psi$$

which indicates that the electric current always flows in the phase planes. This is a necessary result of our assumption that displacement current is negligible.

For waves of the form assumed, the z component of equation (22) and the y component of equation (23) give

$$\frac{E_x}{E_z} = \tan\psi + \frac{\mathrm{i}p^2\epsilon_0}{k^2\sin\psi\cos\psi}.$$

Substituting $p^2/k^2 = V^2S^2 + \mathrm{i}/\epsilon_0$ as found above we have

$$\frac{E_x}{E_z} = -\cot\psi + \frac{\mathrm{i}V^2S^2\epsilon_0}{\sin\psi\cos\psi}.$$

This indicates a component of E_x which is in phase with E_z and lies in the phase planes; it is due to the ohmic resistance of the medium and the current which flows in the phase planes. There is another and much larger component of E_z which is out of phase with E_z; it is an induction electric field due to the varying magnetic field H_y. It may also be shown from equations (22) and (23) that

$$\frac{E_y}{E_z} \simeq -\frac{H_x}{H_y}.$$

There is an additional out-of-phase term which is generally small enough to be neglected.

The Poynting vectors of the two waves rotate in opposite directions on the surface of a cone passing through Oz and $O\xi$ and with axis in the xz plane. When $p \ll \Omega$ the vectors are nearly stationary along Oz and $O\xi$ for the ordinary and extraordinary rays respectively.

(*b*) *The "sound" wave.*—The wave corresponding to the third root of the dispersion equation is found by assuming its velocity to be much less than V. Equations (12) and (17) then give for this wave (remembering $p^2/k^2 \ll V^2$)

$$\frac{H_y}{H_x} = \frac{ip}{\Omega \cos^2\psi}$$

which, together with equations (17) and (11), gives the velocity of the "sound" wave as

$$U\left(\frac{\cos^2\psi - p^2/\Omega^2}{1 - p^2/\Omega^2}\right)^{1/2}.$$

This velocity, as well as other characteristics given below, suggest that this wave is essentially a sound wave, controlled to some extent by the magnetic field. Such a wave can only propagate provided the ion collision frequency is very much greater than the wave frequency. It may be shown that, except in very cold dense gas, this means p/Ω is very small, so that the velocity is approximately $U\cos\psi$.

The component of magnetic field H_y is nearly zero so that the perturbation magnetic field lies along $O\zeta$. The corresponding electric field lies along O_y and the Poynting vector along $O\xi$. The gas velocity is nearly along Oz so that physically the wave is like a sound wave in a gas whose particles may only move in one direction, like beads on a set of smooth parallel wires.

5. *Medium for which $U \gg V$.* (*a*) *The "magnetic" waves.*—In a medium in which the gas pressure is much greater than the "magnetic pressure" one might still anticipate hydromagnetic waves with velocities of the order V. Evidence in this direction is the presence of such waves when $\psi = 0$ (see Section 3). Let us tentatively put $p/k \simeq V$ in equation (11), which then reduces to the approximate form

$$(k^2 + i\epsilon_0 p^2 - i\epsilon_0 V^2 \cos^2\psi k^2) H_x + \frac{\sigma_2}{\sigma_1} \cos^2\psi k^2 H_y = 0. \tag{24}$$

Again write $R = H_y/H_x$ and eliminate H_x and H_y between equations (12) and (24). We find

$$R = \pm 1$$

so that the waves are circularly polarized for all directions of propagation. Substituting $H_y/H_x = \pm i/\cos\psi$ in equation (24) we have the dispersion equation

$$\frac{p^2}{k^2} = V^2 \cos^2\psi \mp \frac{\sigma_2}{\sigma_1 \epsilon_0} \cos\psi + \frac{i}{\epsilon_0}. \tag{25}$$

Substituting $\sigma_1 \epsilon_0 V^2/\sigma_2 = \Omega/p$ this reduces to

$$\frac{p^2}{k^2} = V^2 \cos^2\psi \left(1 \mp \frac{p}{\Omega} \sec\psi + \frac{i}{\epsilon_0 V^2 \cos^2\psi}\right). \tag{26}$$

The waves have velocities $V\cos\psi \mp Vp/2\Omega$ and absorption coefficient

$$\kappa = \frac{p}{2\epsilon_0 V^2 \cos^2\psi}.$$

Thus the assumption that waves with velocity $\simeq V$ were present appears justified.

The components of the gas velocity may be determined from equations (6) to (10), together with the dispersion equation. Combining equations (6) and (26) and remembering that $H_y/H_x = \pm i \sec\psi$, we find

$$v_x = \mp \frac{VH_x}{H_0}\left(1 \mp \frac{p}{\Omega}\sec\psi + \frac{i}{\epsilon_0 V^2 \cos^2\psi}\right)^{-1/2}. \tag{27}$$

Again, from equation (9),

$$v_y = \mp \frac{VH_y}{H_0}\left(1 \mp \frac{p}{\Omega}\sec\psi + \frac{i}{\epsilon_0 V^2 \cos^2\psi}\right)^{-1/2}. \tag{28}$$

It is seen that at frequencies much below the ion gyro-frequency the gas velocity components are approximately

$$v_x = \mp \frac{VH_x}{H_0} \quad \text{and} \quad v_y = \mp \frac{VH_y}{H_0}.$$

The absorption term will be small for any waves which may be propagated to reasonable distances.

From equation (10), neglecting the absorption term,

$$\frac{v_z}{v_x} = -\tan\psi\left\{1 - \frac{V^2}{U^2}\left(1 \mp \frac{p}{\Omega}\sec\psi\right)^{1/2}\right\}^{-1}$$
$$\simeq -\tan\psi.$$

Thus the velocity vector lies approximately in the phase planes. Now introduce $v_p = v_x \sec\psi$, the velocity component in the xz plane. From equations (27) and (28), remembering $H_y/H_x = \pm i \sec\psi$, we find

$$v_y/v_p = \pm i.$$

The motion is circular, the velocity vectors of the ordinary and extraordinary rays rotating in the same directions as those of the magnetic field.

The electric fields of the waves are given by $E_z \simeq 0$ and $E_y/E_x = \pm i \sec\psi$. Again the Poynting vectors rotate on the surface of a cone touching Oz and $O\xi$.

(*b*) *The "sound" wave.*—When the magnetic field is made zero ($V = 0$) a sound wave may be propagated with velocity U. It is, therefore, reasonable to except a sound wave of velocity $\simeq U$ in the presence of a weak magnetic field. This wave may be found by omitting the small absorption terms from equations (11) and (12) and then eliminating H_x and H_y. After some rearranging and substituting Ω/p for $\epsilon_0\sigma_1 V^2/\sigma_2$ we find

$$\frac{p^2}{k^2} - V^2\left(\frac{p^2/k^2 - U^2\cos^2\psi}{p^2/k^2 - U^2}\right) - \frac{p^2}{\Omega^2}V^4\cos^2\psi\left(\frac{p^2}{k^2} - V^2\cos^2\psi\right)^{-1} = 0.$$

Since $p^2/k^2 \simeq U^2 \gg V^2$ and $p/\Omega \gtrsim 1$, the first term is much greater than the last term, which may be simplified by writing U^2 for p^2/k^2. The same is permissible in the first term and the numerator of the second term. Rearranging, we find

$$\frac{p^2}{k^2} = U^2 + V^2\sin^2\psi\left\{1 - \frac{p^2}{\Omega^2}\frac{V^4}{U^4}\cos^2\psi\left(1 + \frac{V^2}{U^2}\cos^2\psi\right)\right\}^{-1}. \tag{29}$$

The factor in the brackets differs very little from unity, so that the wave velocity $\simeq (U^2 + V^2\sin^2\psi)^{1/2}$, which is close to the sound velocity U.

Substituting $p/k = U$ in equation (12) we find

$$\frac{H_y}{H_x} \simeq -\frac{ipV^2}{\Omega U^2},$$

which is very small, so that the magnetic perturbations occur along $O\zeta$.

Substituting $p/k = U$ in equation (10) we find $v_x/v_s = \cot\psi$. Also, from equations (8) and (9) and the fact that H_y/H_x is small we find that v_y/v_s is small. Thus the velocity vector lies very nearly along the wave normal, as would be anticipated. The wave and gas velocities of the three waves which occur when $U \gg V$ are in agreement with the results of Herlofson (**6**) for perfectly conducting media.

The electric perturbations are found to occur along Oy, so that the Poynting vector lies along Oξ.

6. *The physical nature of hydromagnetic waves in gas.* (*a*) *General.*—It is of interest to consider the physical nature of sound and hydromagnetic waves in ionized gas. Sound waves are longitudinal as in non-ionized gas, but owing to the different masses of the electrons and heavy ions, and to their opposite electric charges, the sound waves are accompanied by electric space-charge or plasma waves. The mechanism by which the electron and heavy-ion gases tend to separate and are held together by electric forces has been discussed previously (**13**).

When a magnetic field is present the same effect may be important when $U \geqslant V$, but for the important case when $V \gg U$ the effect may be neglected. Let us then assume only electromagnetic (and not pressure gradient) forces. First assume the medium isotropically conducting so that $\sigma_2/\sigma_1 \to 0$ and consider a plane wave propagated along Ox. Movement of the gas along Ox and across the magnetic field H_0 causes an electric force on the gas $\boldsymbol{v} \times \boldsymbol{H}$ where $\boldsymbol{v}$ is the relative velocity. This field and the resultant current $\boldsymbol{j}$ lie along $-$Oy and so the retarding force $\boldsymbol{j} \times \boldsymbol{H}$ lies along $-$Ox and so opposes motion. Wave motion is thus possible by virtue of electromagnetic forces only.

When the medium is anisotropically conducting, the position is very different. If $\sigma_2 \gg \sigma_1$ the induced electric field along $-$Oy causes current to flow nearly along Ox and a mechanical reaction nearly along $-$Oy at right angles to the direction of motion. The motion of a small element of gas tends to be elliptical rather than longitudinal. The required current along $-$Oy, which causes a mechanical reaction opposing the initial motion along Ox, is provided by the combination of two effects:

(1) The gas develops a velocity component along $-$Oy, thus inducing an electric field in the required direction;

(2) the current along Ox builds up space charge whose potential electric field (along $-$Ox) causes Hall current to flow in the required direction.

(*b*) *Displacement current.*—In order to take into account the space-charge wave which may accompany a hydromagnetic wave, the displacement current term must be introduced into the field equations. This is evident from Maxwell's equation

$$\operatorname{curl} \boldsymbol{H} = 4\pi \boldsymbol{j} + c^{-2} \frac{\partial \boldsymbol{E}}{\partial t}. \tag{30}$$

For a wave travelling in any direction, say Oξ, the space derivatives of $\boldsymbol{H}$ in the plane perpendicular to Oξ must be zero; that is $\operatorname{curl}_\xi \boldsymbol{H} = 0$. Hence if the displacement current term is neglected $j\xi = 0$ and no current may flow to form a space-charge wave.

In the treatment of hydromagnetic waves, as described above, displacement current was neglected. This was done because the equations are made much

more complicated by the introduction of the additional factor. It is usually permissible, provided the wave velocity is much less than the velocity of light. However, in dealing with an anisotropically conducting medium, additional subtleties are introduced and it is desirable to consider the whole question of displacement current. This analysis has been made but is too lengthy to reproduce here. Our simplifying assumption is usually justified at frequencies near or below the gyromagnetic frequency of the heavy ions, but breaks down when V/c approaches unity. It also breaks down for the ordinary ray in the region $\psi \to \pi/2$. In this region the refractive index tends to zero rather than infinity (as shown in Fig. 3) so that such waves may be totally reflected. The extent of the region depends on the magnitude of V/c and is generally small.

The unimportance of displacement current in affecting wave velocity or polarization does not mean that the plasma wave which can occur is also unimportant. This electric wave only requires a very small current, too small to affect appreciably the magnetic field gradients. However, the wave may be important in another way: it may conceivably lead to the spontaneous growth of hydromagnetic waves or, conversely, to the acceleration of ions to relativistic speeds in the manner suggested by Bohm and Gross (**14**) for pure plasma waves. The basis of this mechanism is that a few fast-moving ions are trapped in a potential trough and so travel with the wave. In a medium of slowly varying characteristics the latter may increase or decrease as the wave propagates. This would cause a transfer of energy to the ions or to the wave respectively. It must be remembered that the fast ions must move along the steady magnetic field and so must have velocity $VS \sec \psi$.

Such a mechanism could be important in connection with the generation of cosmic rays or of cosmic radio noise or both.

(*c*) *Hydromagnetic and free-space electromagnetic waves.*—There is a reasonably clear physical picture of a hydromagnetic wave propagated perpendicular to the magnetic field. The electric current flows in the phase plane (the yz plane) and magnetic field and gas density only vary along the wave normal. Thus the wave is a longitudinal pressure wave, like a sound wave, the total pressure comprising that of the magnetic field as well as the gas. There appears to be no correspondingly clear picture of radio, light or other electromagnetic waves in free space. Such a picture is provided by showing that these are extreme examples of hydromagnetic waves.

Without changing the steady magnetic field H_0, let the gas density ρ be reduced towards zero. The hydromagnetic wave velocity is given by $V^2 = H_0^2/4\pi\rho$, but as $\rho \to 0$ the "mass density" of the magnetic field becomes appreciable and must be added to ρ. The energy of the magnetic field is due to a longitudinal tension $H_0^2/8\pi$ in the magnetic lines of force, together with a transverse pressure $H_0^2/8\pi$. The total energy is $H_0^2/4\pi$ and the electromagnetic mass density is $H_0^2/4\pi c^2$. Hence the corrected value of wave velocity is given by

$$V^2 = \frac{H_0^2}{4\pi\{\rho + (H_0^2/4\pi c^2)\}};$$

when $\rho = 0$, $V = c$, and the wave is a hydromagnetic wave propagated through the magnetic lines of force only. The fact that the medium is no longer conducting is not important since displacement current is sufficient to cause the magnetic field gradients and hence the restoring forces opposing such gradients.

The wave velocity when $\rho = 0$ is independent of H_0. This is so because the displacement current is not affected by H_0 (as are currents due to ion movements). Thus H_0 may have any value, including zero, without affecting the wave characteristics, a fact which has been demonstrated experimentally. Radio waves in free space may be regarded as longitudinal pressure waves propagated across a magnetic field whose *mean value is zero*.

(*d*) *Absorption of hydromagnetic waves.*—At first sight it is not clear what value of conductivity should determine the rate of absorption of hydromagnetic waves. However, as Cowling (**12**) has pointed out, although the steady magnetic field controls the motions of the ions, it does not affect their collision rates and so cannot change the Joule losses. It would be anticipated therefore that absorption should be determined by $\sigma_0 (= \sigma_3$ in the gas considered above), the conductivity in the absence of a magnetic field. This is borne out in the preceding investigation in which only σ_3 appears as the imaginary (absorption) term in the dispersion equations.

Division of Radiophysics, C.S.I.R.O.,
Sydney,
Australia :
1955 *March* 23.

References

(1) J. H. Piddington, *M.N.*, **113**, 188, 1953.
(2) H. Alfvén, *Ark. Mat. Astr. Fys.*, **29** B, No. 2, 1942.
(3) H. Alfvén, *Cosmical Electrodynamics*, Oxford, 1950.
(4) C. Walén, *Ark. Mat. Astr. Fys.*, **30** A, No. 15, 1944.
(5) F. de Hoffman and E. Teller, *Phys. Rev.*, **80**, 692, 1950.
(6) N. Herlofson, *Nature, Lond.*, **165**, 1020, 1950.
(7) H. C. van de Hulst, *Problems of Cosmical Aerodynamics*, Chap. VI, Dayton 2, Ohio, Central Air Documents Office, 1951.
(8) N. S. Anderson, *J. Acoust. Soc. Amer.*, **25**, 529, 1953.
(9) E. Astrőm, *Ark. Fys.*, **2**, 443, 1950–51.
(10) V. L. Ginzburg, *Shurnal Expt. Theor. Fiziki U.S.S.R.*, **21**, 788, 1951.
(11) J. H. Piddington, *M.N.*, **114**, 638, 1954.
(12) T. G. Cowling, *The Solar System*, Vol. 1, *The Sun*, edited by G. P. Kuiper, Chicago University Press, 1953.
(13) J. H. Piddington, *M.N.*, **114**, 651, 1954.
(14) D. Bohm and E. P. Gross, *Phys. Rev.*, **74**, 624, 1948.

A SURVEY OF THE INTEGRATED RADIO EMISSION AT A WAVE-LENGTH OF 3·7 M

J. E. Baldwin

(Communicated by M. Ryle)

(Received 1955 August 25)

Summary

A survey of the integrated radio emission between declinations $-28°$ and $+82°$ has been made at a wave-length of 3·7 m using an aerial having beam-widths to half power of $\pm 1° \times \pm 7\frac{1}{2}°$. Except in areas very close to the galactic plane where the resolving power of the aerial is insufficient to show the fine structure of the galactic radiation, the values of brightness temperature are believed to be accurate to within 15 per cent. The results can be interpreted in terms of three symmetrical components extending to progressively greater distances from the plane, as well as some interesting irregular features.

---

1. *Introduction.*—Surveys of the angular distribution of the integrated radio emission have now been made at a number of frequencies in the radio spectrum. At frequencies greater than 300 Mc/s, resolving powers of 2° to 3° have been achieved over large areas of the celestial sphere (**1, 2**); at lower frequencies this detail has not been possible and the surveys have been of two main types:

(1) Surveys covering the whole sky with a resolving power of about $15° \times 15°$ (**3, 4, 5**);

(2) surveys of small areas of sky using a resolving power of about 2° (**6**).

An extended survey has recently been completed by Kraus (**7**) at a frequency of 250 Mc/s using a beam width of $1° \times 8°$. The present survey, covering the area of sky between declinations $-28°$ and $+82°$, used a beam width of $2° \times 15°$ and was made at a low frequency (81·5 Mc/s) where accurate measurements could be made of the absolute radio brightness of the sky in regions away from the galactic plane.

2. *The aerial and receiver.*—The aerial used in this survey, one of the four units of the large Cambridge radio telescope (**8**), was a cylindrical paraboloid 320 ft. × 40 ft., the long axis lying horizontally in an east–west direction. The polar diagram of the aerial in both right ascension and declination was measured in the manner described by Ryle and Hewish (**8**). In this method the recorded deflection is proportional to the voltage polar diagram of the aerial and, by observing an intense radio source, it is possible to determine the polar diagram with great accuracy without appreciable disturbing effects from neighbouring weak sources. In the present measurements the polar diagram could be measured accurately even in directions where the received power was only 0·1 per cent of that in the forward direction. The polar diagrams in right ascension

and declination are shown in Fig. 1. The effect on the observed brightness temperatures of reception of radiation from directions outside the main beam is discussed in Section 4.

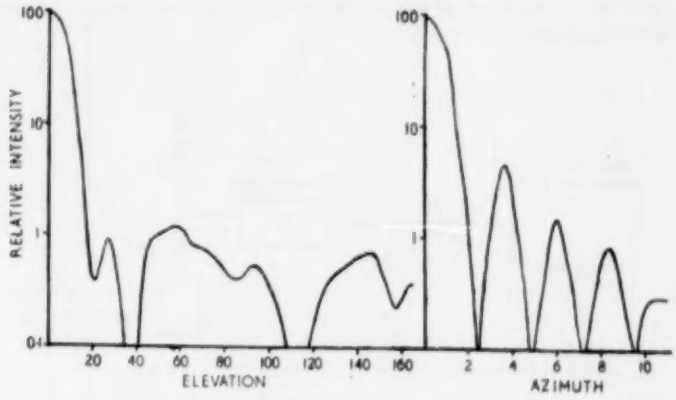


FIG. 1.—*Polar diagram of the aerial system plotted on a logarithmic scale of intensity (a) in declination and (b) in right ascension.*

The signal from the aerial was fed to the receiver through about 800 ft of air-spaced coaxial cable and, in order to overcome the losses introduced by this cable, a preamplifier was installed close to the aerial. The receiver was of the type described by Machin, Ryle and Vonberg (9) in which the signal noise power is compared with the noise power available from a resistance in series with a temperature-limited diode.

3. *Observations and calibrations.*—The aerial beam could be moved in the declination plane in steps of 5°, and 23 settings of the aerial were used at intervals of 5° between declinations +82° and −28°. For declinations lower than −28°, corresponding to elevations of less than 10° above the horizon, the effect of ground reflections became important and these determined the southern limit of the survey. At each setting of the aerial a continuous record of the power received was taken over a period of 24 hours.

It is convenient to express the noise power received in terms of an effective temperature. The brightness temperature of the sky, T_b, is defined as the temperature of a black body for which the brightness of the thermal radiation at a particular frequency would be the same as that observed. The available noise power corresponding to an effective temperature T_b is then $kT_b\delta f$ where δf is the band-width of the receiving system.

With the arrangement used in the present survey the effective temperature T_r of the signal reaching the receiver is

$$T_r = \beta g(\alpha T_b + T_n + (1-\alpha)T_0) + (1-\beta)T_0,$$

where α is the loss in the cable between the aerial and the preamplifier, β is the loss in the cable between the preamplifier and the receiver, g is the gain of the preamplifier, T_0 is the ambient temperature, and T_n is the effective noise temperature of the preamplifier.

The observations provide a 24-hour record of T_r and, in order to derive the values of T_b in different directions, it is necessary to eliminate the quantities α, β, g, and T_n. Daily calibrations of the sensitivity and zero-level of the system were made by replacing the aerial cable at the input of the preamplifier by a known source of noise power, provided by a resistance in series with a temperature-limited diode, as shown in Fig. 2.

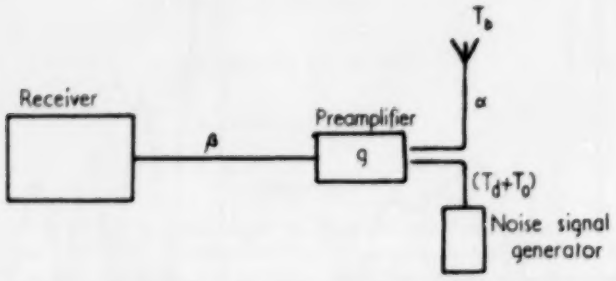


FIG. 2.—*Arrangement used for measuring the sensitivity of the system.*

The values of T_r produced (1) with a given diode current corresponding to an equivalent temperature $T_d + T_0$ and (2) with no diode current, corresponding to an equivalent temperature T_0, are:

$$T_{r1} = \beta g(T_d + T_0 + T_n) + (1 - \beta)T_0$$
$$T_{r2} = \beta g(T_0 + T_n) + (1 - \beta)T_0.$$

As the value of g was very large, the term $(1 - \beta)T_0$ could be neglected and hence the true brightness temperatures T_b could be obtained in terms of T_d, T_0 and α. The attenuation α of the cable between the aerial and the preamplifier was determined by a separate experiment.

Since both T_n and g depend on the impedance of the source and since neither the noise signal generator nor the aerial were perfectly matched to the coaxial feeder, corrections were made by inserting different lengths of cable between both the signal generator and preamplifier, and the aerial and preamplifier.

4. *The method of analysis.*—Using the methods of calibration described above, the observed values of T_r obtained throughout 24 hours of right ascension on successive 5° strips of declination could be converted into apparent brightness temperature. In making this analysis, all discrete sources which produced a contribution to the brightness temperature greater than 40 deg. K were noted from the Cambridge survey of radio stars (**10**) and their effects were subtracted from the records.

In combining the values of brightness temperature to form a map, no corrections were made for the finite width of the reception pattern. Previous surveys (**3**, **4**, **5**) have used such corrections and a general account of their application has been made recently by Bracewell and Roberts (**11**); they were not used in the present survey for the following reasons. In regions away from the galactic plane the resolving power appeared to be sufficient to resolve the main features of the background whilst, in regions near the plane, the presence of the narrow belt of intense radio emission lying along the galactic plane would have produced spurious subsidiary maxima if this method of aerial beam correction

had been applied. An indication of the true contours in the latter regions has been obtained using a system of considerably greater resolving power (**12**); the results are shown in Fig. 3.

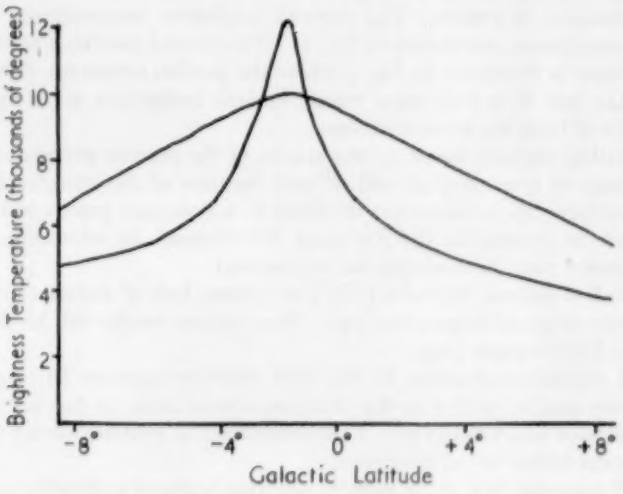


FIG. 3.—*The variation of brightness temperature with galactic latitude at* $l = 353°$.
Upper curve - - - present observations
Lower curve - - Scheuer and Ryle (**12**).

In addition to the effects caused by the limited resolving power it is also important to examine the magnitude of the errors caused by reception of radiation through subsidiary maxima in the reception pattern. With the present polar diagram (Fig. 1) it can be shown that, in the areas where the brightness temperature is small, the observed temperature might differ from the true values by not more than 50 deg. K. In a similar way the highest temperatures observed ($> 10^4$ deg. K) will be lower than the true ones by about 5 to 10 per cent. As the first effect is small and the second is only appreciable in those areas of sky where the aerial beam is larger than the fine structure of the galactic radiation, both effects have been neglected in the quoted temperatures.

The sources of error arising in the determination of the brightness temperatures are listed below:

(1) The effects of reception of radiation through subsidiary maxima of the main beam. This has been discussed above.

(2) The loss in the cable between the aerial and preamplifier. This quantity was measured to an accuracy of 3 per cent.

(3) Drift in the recorder sensitivity or zero level, due to variations in the preamplifier or the aerial impedance. Errors arising from these causes are estimated to be not greater than 3 per cent.

(4) The reading accuracy of the records. This was not found to be a significant limitation on the accuracy of the quoted temperatures.

(5) The fundamental accuracy of the standard of noise power used in calibrating the system. The diode noise source was calibrated by Adgie (**13**) against a thermal source with an accuracy of 5 per cent.

It is estimated that the absolute accuracy of the quoted temperatures is 15 per cent, except in regions near the galactic plane where the resolving power is insufficient, whilst the accuracy of the ratio of the temperatures in different parts of the sky is probably about 5 per cent.

5. *Discussion of results.*—The derived brightness temperatures, plotted in galactic coordinates, are shown in Fig. 4. The limited resolving power near the galactic plane is illustrated in Fig. 3 where the profiles across the plane obtained by Scheuer and Ryle (**12**) using interferometric techniques are compared with those derived from the present survey.

A detailed analysis, based on the results of the present survey, has revealed the presence of a number of well defined features of the integrated radiation. These features will be discussed in detail in subsequent papers but it may be useful here to summarize the five main distributions by which the integrated radio emission may conveniently be represented:

(1) At low galactic latitudes ($<3°$) an intense belt of emission is detectable over a wide range of longitudes (**12**). More recent results will be described in detail in a further paper (**14**).

(2) A significant fraction of the total emission appears to originate in a distribution similar to that of the common optical stars, as has been discussed by Westerhout and Oort (**15**). This component is responsible for the bulk of the radiation within 20° of the plane.

(3) The greater part of the galactic radiation is due to a distribution of almost spherical form, as suggested by Shklovsky (**16**), having a radius of 11 to 16 kps. Evidence for a similar population in the Andromeda nebula has already been described (**17**).

(4) Extragalactic sources contribute an isotropic component to the integrated radiation. A discussion of the possible magnitude of this component will be given in a later paper.

(5) In addition to the above components of the integrated radiation, there are a number of well defined features which are too extended to be classed as individual radio sources but which cannot be fitted into any symmetrical model. One of the most marked of these features, which has been observed by Kraus and Ko (**18**) and by Hanbury Brown and Hazard (**19**) has been interpreted as the radio emission from the local supergalaxy. The possible nature of this and other irregular features will be discussed in a later paper.

This work was carried out as part of a programme of research supported in the Cavendish Laboratory by the Department of Scientific and Industrial Research, to which the author is indebted for a maintenance allowance. The author wishes to express his thanks to Mr M Ryle for his continual help and advice.

Cavendish Laboratory,
Cambridge:
1955 *August* 22.

References

(1) G. Reber, *Proc. Inst. Rad. Eng.*, **36**, 1215, 1948.
(2) J. F. Denisse, E. Leroux and J. L. Steinberg, *Comptes Rendus*, **240**, 278, 1955.
(3) J. S. Hey, J. W. Parsons and S. J. Phillips, *Proc. Roy. Soc.*, A **192**, 425, 1948.
(4) J. G. Bolton and K. C. Westfold, *Aust. J. Sci. Res.*, A3, 19, 1950.
(5) C. W. Allen and C. S. Gum, *Aust. J. Sci. Res.*, A **3**, 224, 1950.
(6) R. Hanbury Brown and C. Hazard, *M.N.*, **113**, 109, 1953.

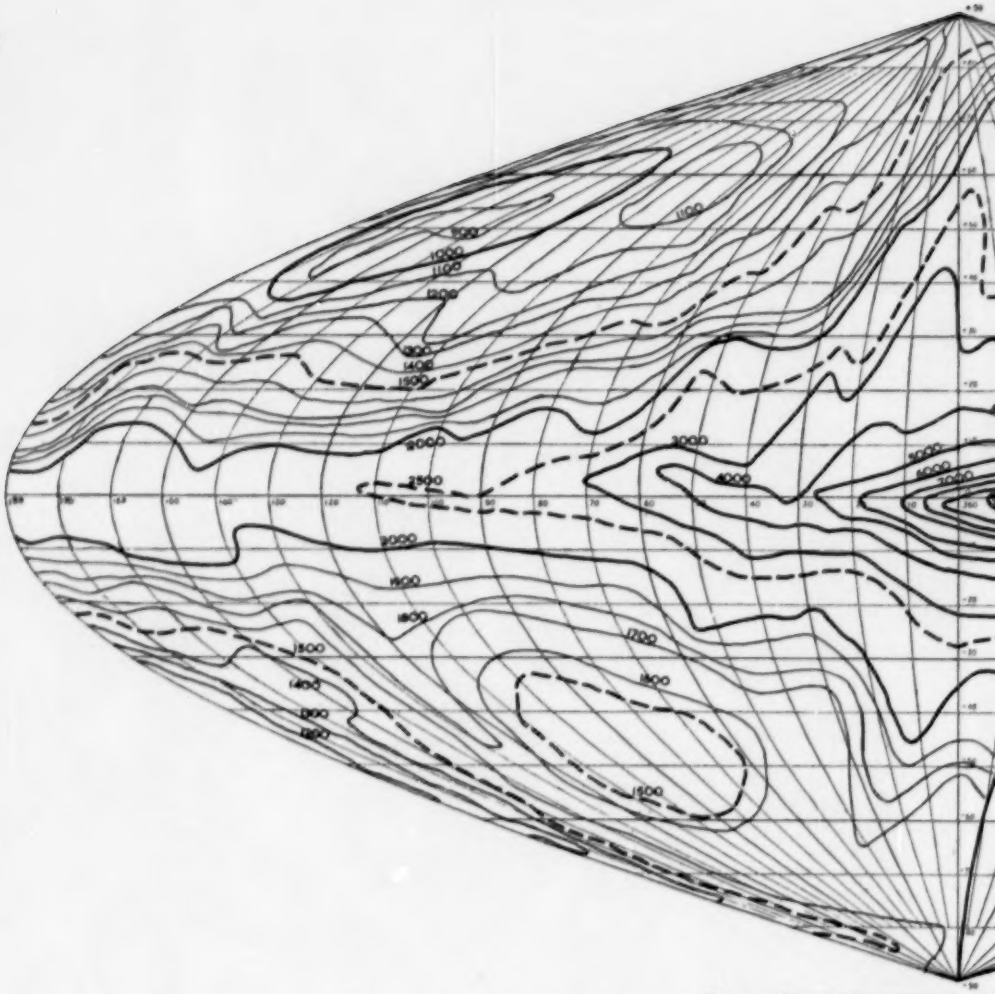
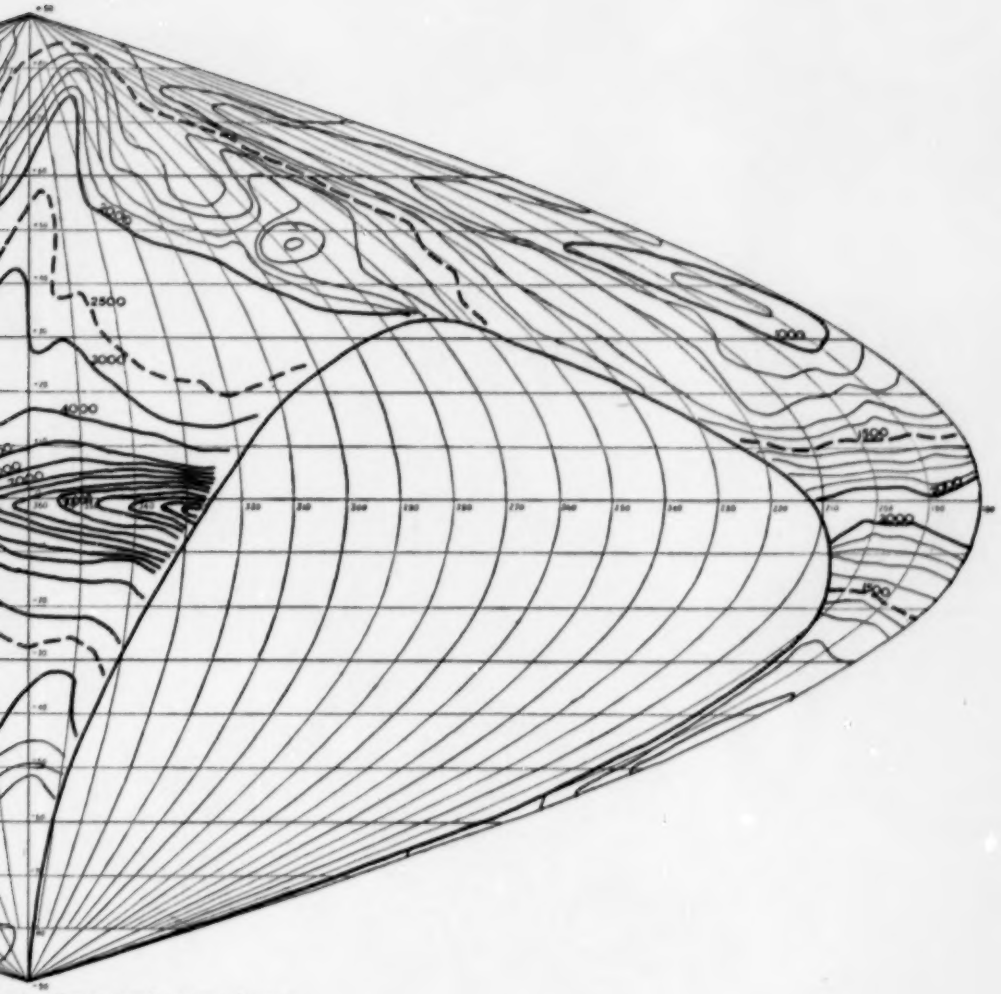


FIG. 4.—*Map of the integrated radi*
——— *intervals of*
– – – – *intervals of*
——— *intervals of*



ted radiation in galactic coordinates.
rvals of 100 deg. K
rvals of 500 deg. K
rvals of 1000 deg. K.

(7) J. D. Kraus and H. C. Ko, *Nature*. **175**, 159, 1955.
(8) M. Ryle and A. Hewish, *Mem. R.A.S.*, **67**, Part 3, 1956.
(9) K. E. Machin, M. Ryle and D. D. Vonberg, *Proc. I.E.E.*, **99**, 127, 1952.
(10) J. R. Shakeshaft, M. Ryle, J. E. Baldwin, B. Elsmore and J. H. Thomson, *Mem. R.A.S.*, **67**, Part 3, 1956.
(11) R. N. Bracewell and J. A. Roberts, *Aust. J. Phys.*, **7**, 615, 1954.
(12) P. A. G. Scheuer and M. Ryle, *M.N.*, **113**, 3, 1953.
(13) R. Adgie (in preparation)
(14) J. E. Baldwin and P. A. G. Scheuer (in preparation)
(15) G. Westerhout and J. H. Oort, *B.A.N.*, **11**, 323, 1951.
(16) I.S. Shklovsky, *Ast. Zh.*, **29**, 418, 1952.
(17) J. E. Baldwin, *Nature*, **174**, 320, 1954.
(18) J. D. Kraus and H. C. Ko, *Nature*, **172**, 538, 1953.
(19) R. Hanbury Brown and C. Hazard, *Nature*, **172**, 997, 1953.

THE DISTRIBUTION OF THE GALACTIC RADIO EMISSION

J. E. Baldwin

(Communicated by M. Ryle)

(Received 1955 August 25)

Summary

An analysis is made of the distribution of the integrated radio emission, and it is shown that most of the emission is galactic in origin and arises from a spherical distribution having a radius greater than 10 kps; the emission per unit volume throughout this spherical region is remarkably constant. A further distribution which accounts for the greater part of the radiation at galactic latitudes less than 30° is found to be consistent with the radio model of the Galaxy proposed by Westerhout and Oort. A comparison of the Galaxy with M31 shows that, whilst the extents of the spherical components in the two galaxies are very similar, the emission per unit volume is about six times greater in the Galaxy than in M31.

---

1. *Introduction.*—A survey of the galactic radio emission by Bolton and Westfold (**1**) at a wave-length of 3·0 m was used by Westerhout and Oort (**2**) as the experimental basis for a radio model of the Galaxy. Their model was based on the supposition that the sources responsible for the background radiation were distributed in space in the same way as the distribution of mass in our Galaxy, as deduced from optical studies by Oort and van Woerkom (**3**). Using this distribution they found that the derived contours of radio brightness showed good general agreement with those observed except in two respects:

(1) The observed brightness temperatures in directions near the anti-centre were considerably greater than the calculated values: Westerhout and Oort associated these high temperatures with emission from the outer spiral structure of the Galaxy.

(2) Over the whole sky the observed temperatures showed a fairly uniform excess of about 600 deg. K (at a wave-length of 3·0 m) above the calculated temperatures. Westerhout and Oort suggested that this emission might originate in extragalactic sources or in some widely extended population of galactic sources.

Later observations of the background radiation using considerably higher resolving powers (**4**) have shown in addition the existence of an intense band of radio emission about 2° in width lying along the galactic equator.

A new extensive survey of the background radiation, at a wave-length of 3·7 m, has now been completed (**5**) as well as further more detailed interferometric observations. The latter observations and the nature of the sources contributing to the concentrated belt of emission will be discussed in a later paper (**6**), where it is shown that these sources contribute only about 5 per cent of the total galactic radiation. The present paper is concerned with the origin of the radiation outside this narrow belt. All brightness temperatures quoted will refer to the wave-length of the new survey, 3·7 m.

2. *The radiation at high galactic latitudes.*—According to the model proposed by Westerhout and Oort, the areas of minimum brightness in the sky should lie within a few degrees of the galactic poles. In fact, as pointed out by Wyatt (**7**), these regions lie at galactic latitudes of about $\pm 45^\circ$ between longitudes 100° and 210°. The positions of these minima indicate that an appreciable fraction of the radiation from high galactic latitudes may be galactic in origin.

As a guide to the most appropriate model to adopt it is helpful to consider the results of observations of the distribution of radio brightness across the Andromeda nebula. It was shown (**8**) that the radio emission extended to far greater distances from the centre, both in the plane and perpendicular to the plane of the Galaxy, than would have been expected from the Westerhout and Oort model. The results were interpreted in terms of a simple model consisting of (*a*) a sphere of radius 10 kps, having a uniform emission per unit volume, which contributed about two-thirds of the total radio emission from the nebula, and (*b*) a population similar to that proposed by Westerhout and Oort contributing the remaining one-third.

Shklovsky (**9**) originally put forward the idea that the radiation from our own Galaxy could be interpreted in similar terms. It is important to see whether the present observations support this suggestion. In directions close to the galactic plane the radiation from a spherical population and one similar to that of the Westerhout and Oort model will be superimposed, but at galactic latitudes greater than 40° the contribution from the latter population is < 100 deg. K, i.e. less than one-tenth of the total emission observed in these directions. An attempt will therefore be made to see how far it is possible to interpret the observed distribution for $b > 40^\circ$ in terms of an extended ellipsoidal model. It will first be assumed that there is no contribution from extragalactic sources; the effects on the model of different amounts of extragalactic radiation will then be examined.

Model 1

The first model considered consists of a spherical distribution having a uniform emission per unit volume. If it is assumed to be optically thin for radiation of this wave-length (3·7 m) then, in any direction, the observed brightness is proportional to the length of the line SN in Fig. 1. A range of models may be constructed by varying both the radius of the sphere, R, and the emission per unit volume, σ. For the particular latitude of $b = +50^\circ$, the observed temperatures as a function of longitude are given in the form of a polar plot in Fig. 2; with a spherical model, the curves for a given latitude on this projection are circles. The contours predicted for models having radii R of 12 kps, 16 kps and 20 kps are shown in the figure, the distance SC having been assumed to be 8 kps. In each case the value of σ has been adjusted to give the best fit to the experimental curve. A change in the value of σ corresponds to a proportional change in the brightness temperatures in all directions. It may be seen that the sphere of radius 16 kps gives the best fit to the observations.

Using the values $R = 16$ kps and $\sigma = 1{\cdot}8 \times 10^{8}$ watts ster^{-1} (c/s)$^{-1}$ ps^{-3} determined from this figure, the expected brightness distributions at latitudes of $+30^\circ$, $+40^\circ$, $+60^\circ$, and $+70^\circ$ have been computed and are shown in Fig. 3 together with the observed temperatures. The agreement is good over most longitudes, appreciable departures occurring only in the neighbourhood of the arm of radiation at $l = 0^\circ$; in the curves for $b = 30^\circ$ a discrepancy due to the contribution of the Westerhout and Oort population is to be expected.

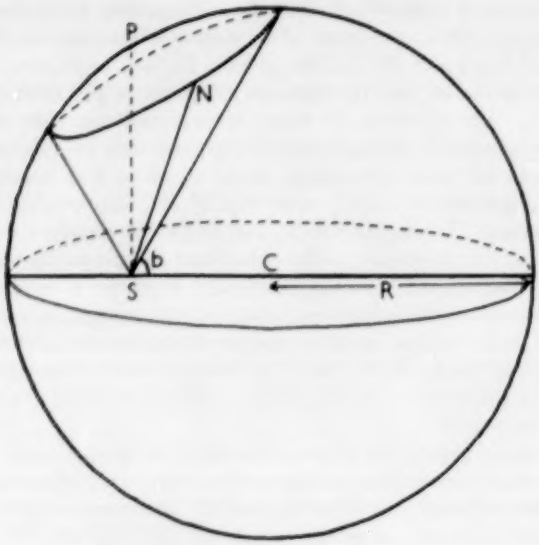


FIG. 1.—*Diagram showing geometrical relationships.*

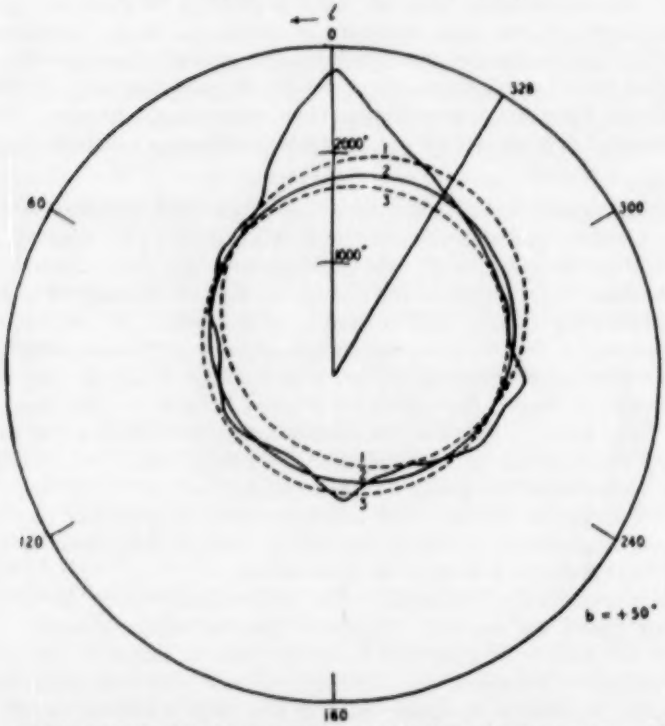


FIG. 2.—*The variation of brightness temperature with galactic longitude at* $b = +50°$. *The three circles are the calculated distributions for models having values of R of* (1) 12 *kps* (2) 16 *kps and* (3) 20 *kps respectively.*

So far it has been assumed that the spherical component of the galactic radiation has a uniform emission per unit volume. Although this assumption leads to good agreement between the observed and calculated distributions of temperature, there are two regions in which appreciable variations in the value of σ would be undetectable.

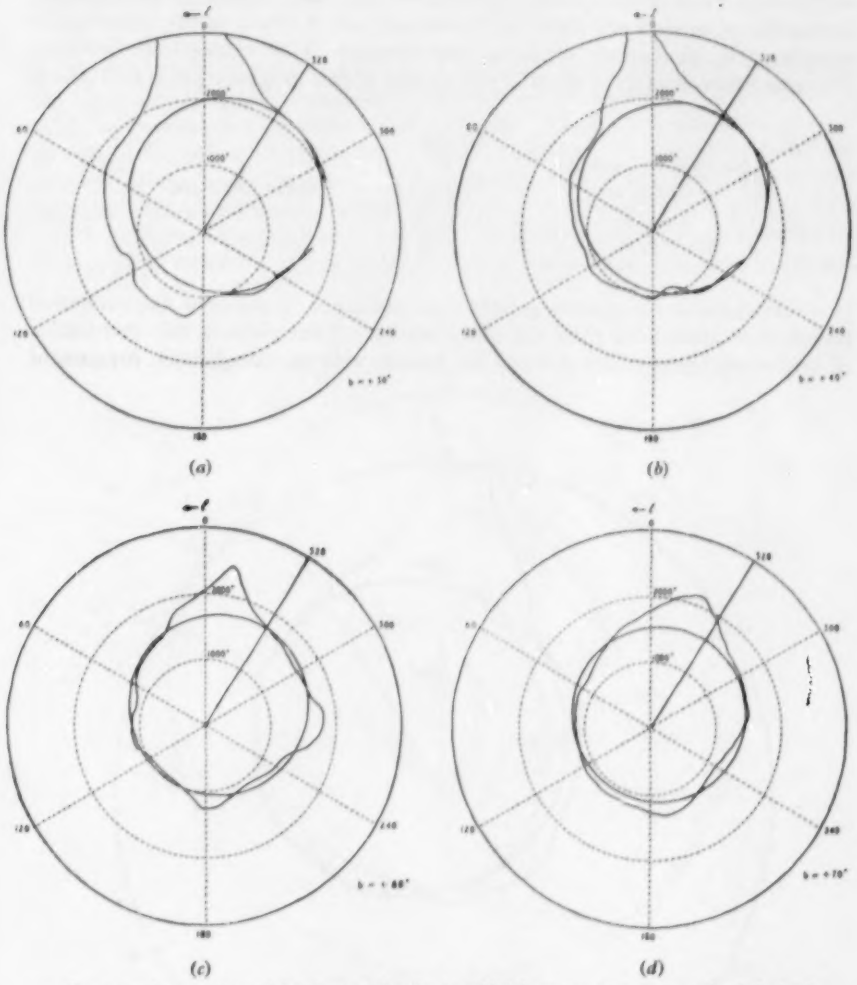


FIG. 3.—*A comparison of the observed distribution of brightness temperature at galactic latitudes of +30°, +40°, +60°, and +70° with the predicted curves for a spherical model having R = 16 kps and* $\sigma = 1{\cdot}8 \times 10^{8}$ *watts ster*$^{-1}(c/s)^{-1}$ *ps.*$^{-3}$

(1) The region within 4 kps of the centre of the Galaxy. Observations are of necessity confined to galactic latitudes $<30°$ where it is difficult to separate the radiation originating in the spherical population from that of the Westerhout and Oort population; the value of σ here may be significantly different from that in the range 4–10 kps from the centre.

(2) The space beyond about 10 kps from the centre of the Galaxy. The difficulty of the interpretation of this region is that a spherical shell of sources whose radius is appreciably greater than SC gives a distribution of brightness over the sky which is very nearly isotropic as seen from the neighbourhood of the Sun. It is thus not possible to determine the radial distribution of the sources in the outer parts of the sphere; the sources could for example be extragalactic. A number of models can therefore be derived, all of which fit the observations equally well, containing different contributions from extragalactic sources. The essential constants of three of the models in this range are given in Table I.

TABLE I

Assumed extragalactic contribution	$\bar{R}$ (kps)	σ (watts $ster^{-1}(c/s)^{-1}$ ps^{-2})
0 deg. K	16	$1{\cdot}8\times10^{5}$
250 deg. K	13·5	$1{\cdot}8\times10^{5}$
500 deg. K	11	$1{\cdot}8\times10^{5}$

If an extragalactic component greater than 500 deg. K is assumed, the computed curves show deviations from the observations. Polar plots of the distribution of brightness temperature derived for models with an extragalactic component

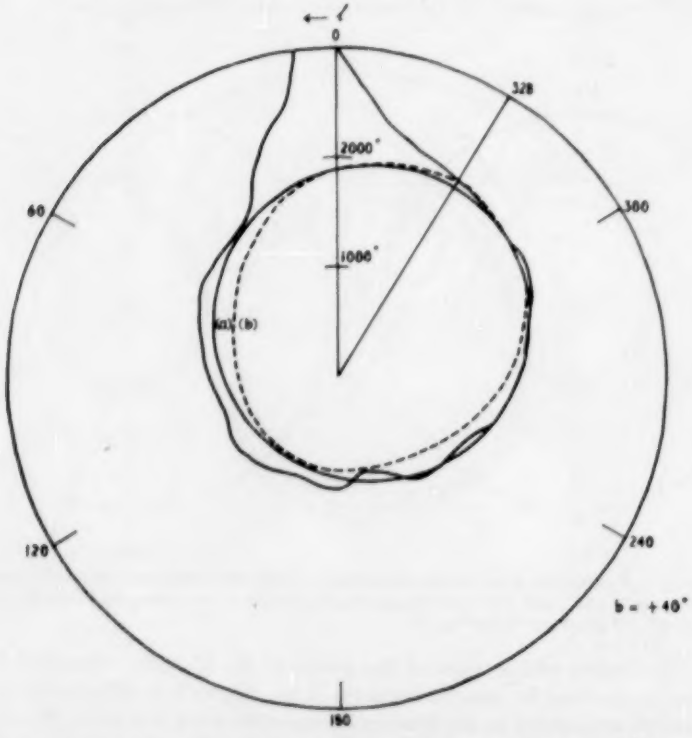


FIG. 4.—*Computed curves for models having different contributions from extragalactic sources: (a) ⩽ 500 deg. K, (b) 700 deg. K.*

of (*a*) $\leqslant$500 deg. K and (*b*) 700 deg. K and the corresponding values of R are shown in Fig. 4. It can be seen that an extragalactic component $>$500 deg. K is not compatible with the observations.

Before considering the implications of this result it is important to see whether other distributions of the extended galactic component could provide an equally good fit with the observations.

Model 2

In this model the distribution is in the form of an ellipsoid of revolution, its minor axis lying perpendicular to the plane of the Galaxy. As in Model I, a uniform emission per unit volume is assumed. For a particular axial ratio of the ellipsoid, values of the major axis and emission per unit volume can be chosen so that the theoretical and observed curves fit closely at any one latitude. Three ellipsoids having axial ratios of 1·0, 0·7 and 0·5 have been fitted to the observations at $b = +40°$. A comparison of the corresponding curves at $b = +60°$ is shown in Fig. 5. By examining the curves for different

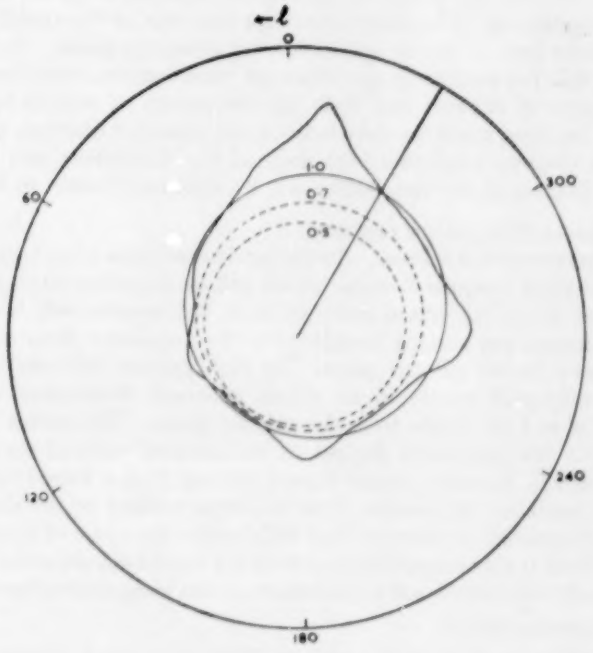


FIG. 5.—*Calculated distribution of brightness temperature for models with ellipsoids having different axial ratios.*

latitudes and the effect of assuming different values of the extragalactic component, a restricted range of homogeneous ellipsoidal models which will fit the observations can be constructed. They are given in Table II. As before, the extragalactic radiation cannot exceed 500 deg. K. Before considering how

recent estimates of the integrated radiation from extragalactic sources may be used to narrow the range of possible models, we shall discuss the observed distribution of emission at low galactic latitudes.

TABLE II

Assumed extragalactic	Major axis (kps)	Axial ratio
0 deg. K	16	>0·7
250 deg. K	13·5	>0·6
500 deg. K	11	>0·5

3. *The emission at low galactic latitudes.*—At galactic latitudes less than 30° the effects of less extended components of the galactic radio emission are superimposed on the contributions from the 'spherical' and extragalactic components. In Fig. 6 these latter have been subtracted from the observed contours. For latitudes less than 30° the distribution of brightness agrees reasonably well with that calculated on the basis of the model of Westerhout and Oort. In Fig. 7 the observed variation of brightness temperature with l (at $b=0°$) and with b (at $l=340°$) are compared with those expected from the model. The observed distribution in latitude is narrower than that of the model; this is due to the narrow belt of intense emission lying along the plane. In Fig. 7(*b*) the effects of this population on the observed temperatures, calculated from the measurements of Scheuer and Ryle (**4**), are shown as vertical lines. When allowance has been made for this emission, the observed contours agree both in shape and absolute brightness with those of the Westerhout and Oort model. Irregular features of the distribution will be discussed further in Section 4(*b*).

4. *A model of the galactic radiation*

(*a*) *The symmetrical features.*—In the previous sections it has been shown that, apart from some irregular features which will be described later, the observed distribution of the integrated radio emission can conveniently be represented by two extended populations in addition to the population showing very strong concentration to the galactic plane. At high galactic latitudes the observed emission originates mainly in an almost spherical distribution extending to distances of at least 10 kps from the galactic plane. The extent and possible ellipticity of this population depend on the assumed value of the extragalactic emission which, however, cannot exceed 500 deg. K at a wave-length of 3·7 m. It is now necessary to examine other evidence relating to the emission from extragalactic sources in order to limit still further the range of possible models.

Shakeshaft (**10**) has suggested that there are three main classes of extragalactic object which make significant contributions to the integrated emission:

(i) normal galaxies;

(ii) galaxies such as NGC 1275 and NGC 2623 which appear to represent encounters between two nebulae and which have a radio luminosity some 5×10^2 times that of our own Galaxy or M 31;

(iii) sources similar to that in Cygnus, which has been interpreted by Baade and Minkowski (**11**) as a collision between two late-type spirals: the radio luminosity of this source is about 5×10^5 times that of our Galaxy.

Shakeshaft also showed that the value of the integrated radiation depends on the cosmological theory assumed.

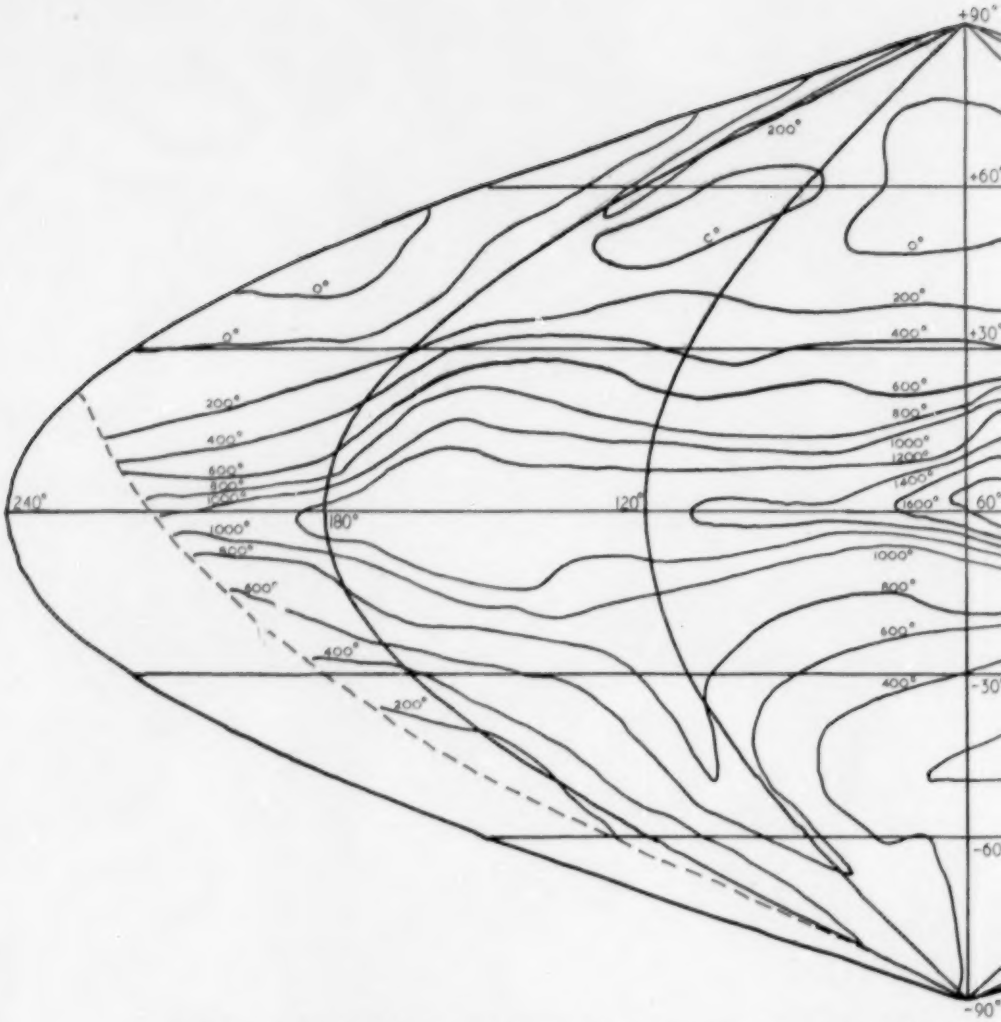
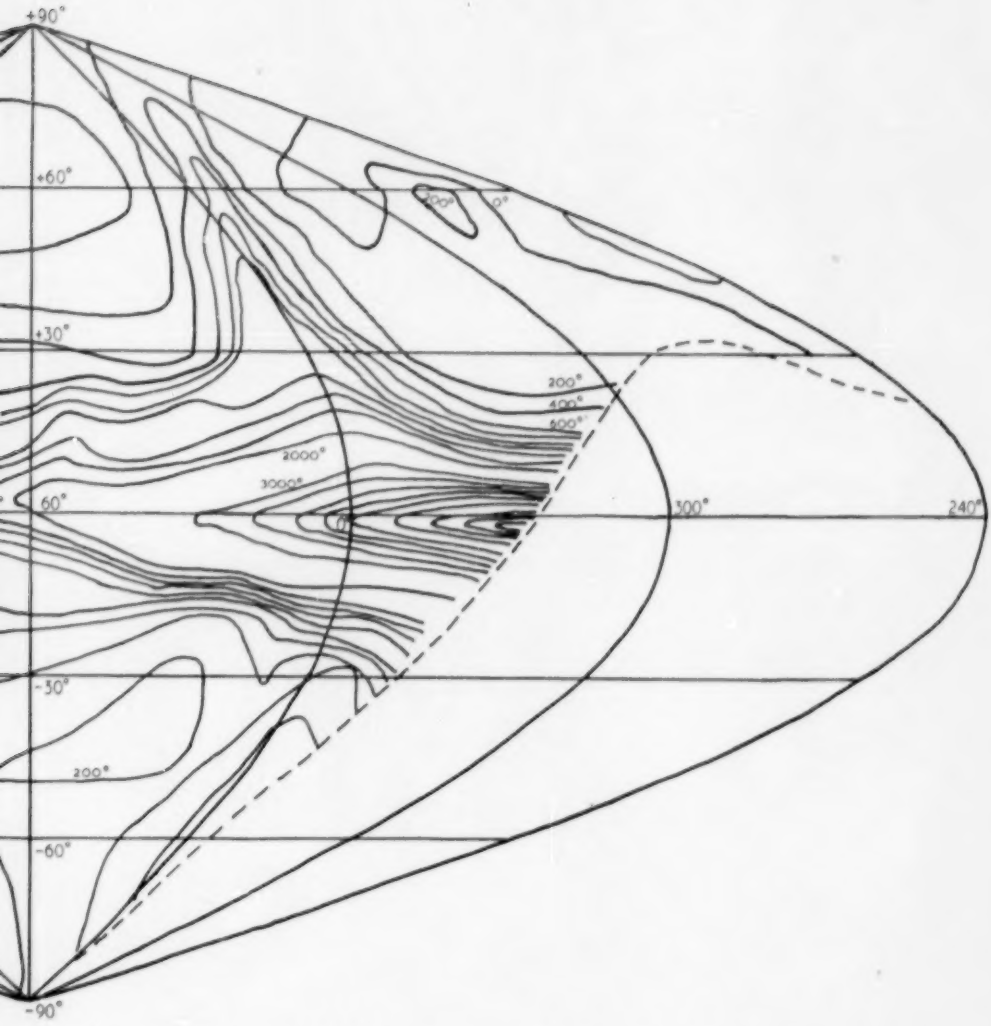


FIG. 6.—*The distribution of brightness temperature after subtracting the contribution from the spheric*
of 2000 deg. K and at intervals of 1



spherical component. The contours of brightness are at intervals of 200 deg. K up to temperatures als of 1000 deg. K above this value.

In a recent paper (**12**) evidence was presented suggesting that the majority of the observed radio stars belong to class (iii). It was also shown that the results of the number-magnitude plot were difficult to explain in terms of steady-state theories of cosmology. A full discussion of these new data will be given elsewhere, but a preliminary estimate suggests that extragalactic sources must make a contribution of at least 200 deg. K to the total radiation. Since it has already been shown that the observed distribution cannot be explained in terms of a symmetrical galactic model if the extragalactic component exceeds 500 deg. K, it is now possible to define the extended component of the galactic emission within rather small limits.

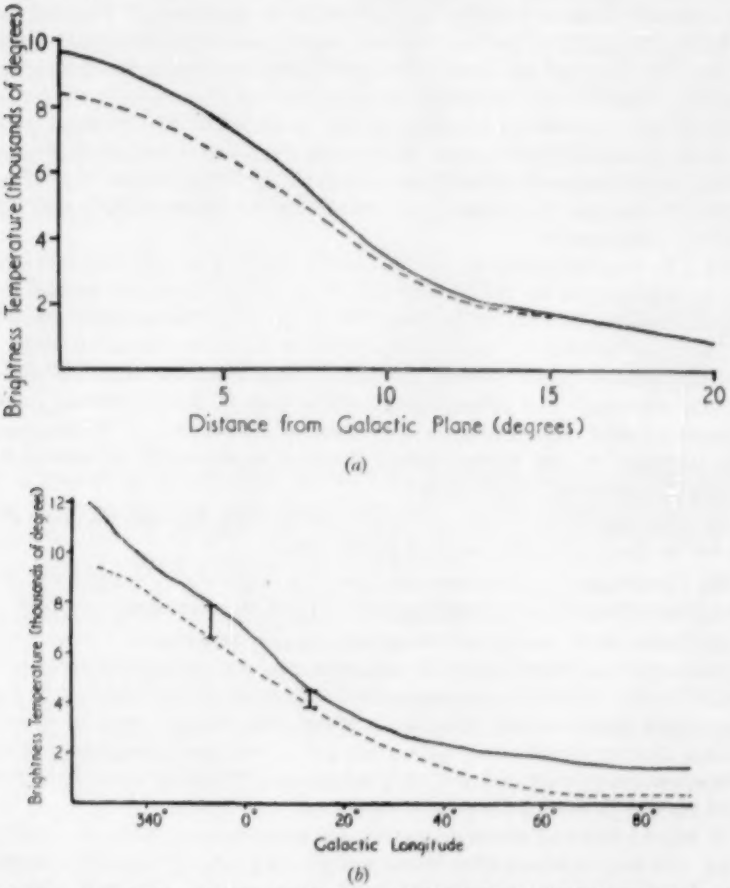


FIG. 7.—*A comparison between the brightness distribution in (a) latitude and (b) longitude derived from Fig. 6 (full line), and that of the Westerhout and Oort model (dotted line). The contributions to the aerial temperature to be expected from the very narrow belt of emission lying along the plane are shown as vertical lines in Fig. 7(b).*

The following model of the galactic radio emission is therefore proposed:

(i) An ellipsoidal distribution having a major axis of 11–14·5 kps and an axial ratio greater than 0·5 (for $R = 11$ kps) or greater than 0·6 (for $R = 14{\cdot}5$ kps).

The emission per unit volume shows no marked variation with radial distance from the galactic centre, and has a value at a wave-length of 3·7 m of $1{\cdot}8 \times 10^{8}$ watts ster^{-1} (c/s)$^{-1}$ ps^{-3}.

(ii) A distribution similar to the model of Westerhout and Oort. The emission per unit volume agrees closely with that derived by Westerhout and Oort, a spectral variation as $\lambda^{0{\cdot}5}$ being assumed in comparing the values at 3·0 m and at 3·7 m.

(iii) A disk-shaped distribution concentrated within about 150 ps of the galactic plane. This will be discussed in detail in a later paper (**6**).

The distinction between (i) and (ii) may be purely one of descriptive convenience; there is no evidence to show that the emission from (i) and (ii) has different origins. Observations have been made by Adgie (**13**) to determine accurately the spectral variation in directions (1) close to the galactic plane which, when corrected for the effects of distribution (iii), corresponds mainly to the radiation from (ii) and (2) near the galactic poles where the emission is mainly due to the spherical distribution. His results show that in the range 38–175 Mc/s the brightness temperature has a variation (1) $\lambda^{2{\cdot}51 \pm 0{\cdot}10}$ and (2) $\lambda^{2{\cdot}41 \pm 0{\cdot}10}$. This result indicates that the sources responsible for the emission in (i) and (ii) may be of the same nature.

(*b*) *The irregular features.*—Although the majority of the observed emission may be represented by the models described above, there are several marked irregular features that cannot be described in terms of a symmetrical model.

The most notable of these is the band of emission extending along $l = 0°$ from the galactic plane towards the pole. This has been observed in several previous surveys. The present observations suggest that it extends round the complete celestial sphere, roughly in the form of a great circle. Whilst it appears to be unrelated to any known optical object, it is reasonable to assume it to be galactic in origin for three reasons:

(*a*) The brightness of the band increases towards the galactic plane at $l = 0°$ and has its maximum value on or near the plane.

(*b*) The brightness temperatures observed in parts of the band are as great as or greater than the integrated brightness from all extragalactic sources.

(*c*) In a recent survey of radio stars (**14**) a remarkable concentration of 6 intense stars was found between longitudes 115° and 150° and between latitudes 0° and −12°. This area corresponds closely with the intersection of the belt of enhanced emission with the galactic plane, and although there is no evidence to relate the two phenomena, the agreement in position is suggestive that these radio stars are galactic and that they arise from the interaction of the material in the band with that in the galactic plane.

A second band of emission having its maximum intensity at $l = 273°$ and $b = 54°$ has been attributed by Kraus and Ko (**15**) and by Hanbury Brown and Hazard (**16**) to radiation from the local super-galaxy. The new observations have indicated that this feature has a maximum brightness temperature (at a wave-length of 3·7 m) of about 400 deg. K, a figure comparable with the total emission from extra-galactic sources; its interpretation in terms of emission from the supergalaxy therefore presents considerable difficulties and a galactic origin of this feature seems possible. A fuller discussion of this problem will be given in a subsequent paper.

5. *Comparison with the Andromeda nebula.*—Observations of the Andromeda nebula (**8**) have shown that in this case also most of the radiation originates in an extended spherical distribution and that less than one-third of the total radiation is contributed by a component similar to that of the Westerhout and Oort model. The distribution of brightness across the nebula is consistent with a spherical distribution having a uniform emission per unit volume and an angular radius of 100′ of arc. Recent measurements have given the distance of M 31 as 610 kps (**17**); using this figure the radius of the spherical distribution is 18 kps.

In Table III a comparison is made between the radio emission from the Andromeda nebula and from the two limiting models of the spherical distribution in the Galaxy. The total emission from the Westerhout and Oort population of sources for the Galaxy is also given; it can be seen to make a relatively small contribution to the total radiation. The values of σ given in the last column refer to the emission per unit volume in the spherical distribution. The two figures quoted for the Andromeda nebula refer to the cases in which (i) the spherical distribution contributes only two-thirds of the total emission (this was the lower limit set by the observations) and (ii) the spherical distribution is responsible for all the radiation.

TABLE III

	$\bar{R}$ (kps)	Total emission watts ster^{-1} (c/s)$^{-1}$	σ watts ster^{-1} (c/s)$^{-1}$ ps^{-3}
M31	18	$0{\cdot}8\times10^{21}$	$0{\cdot}24-0{\cdot}35\times10^{8}$
Spherical models	14·5	$2{\cdot}3\times10^{21}$	$1{\cdot}8\times10^{8}$
of the Galaxy	11	$1{\cdot}0\times10^{21}$	$1{\cdot}8\times10^{8}$
W. and Oort distribution		$0{\cdot}3\times10^{21}$	

It is interesting to note that, although the limitations of the present analysis lead to an uncertainty of more than a factor of 2 in the total emission from the Galaxy, the derived emission per unit volume is much more accurately known and is some six times greater than the corresponding figure for M31.

6. *Conclusion.*—It has been found possible to interpret most of the galactic radio emission in terms of a spherical distribution having a radius of between 11 and 14·5 kps: the emission per unit volume is remarkably constant and between radii of 4 and 10 kps cannot vary by more than a factor of 2. An attempt to explain such a distribution meets with considerable theoretical difficulties; some of the problems involved and the possible origin of the radiation are discussed in detail by Burbidge (**18**).

This work was carried out as part of a programme of research supported in the Cavendish Laboratory by the Department of Scientific and Industrial Research, to which the author is indebted for a maintenance allowance. He also wishes to thank Mr M. Ryle for valuable advice in this work.

Cavendish Laboratory,
Cambridge:
1955 *August* 22.

References

(1) J. G. Bolton and K. C. Westfold, *Aust. J. Sci. Res.*, A **3**, 19, 1950.
(2) G. Westerhout and J. H. Oort, *B.A.N.*, **11**, 323, 1951.
(3) J. H. Oort and A. J. J. van Woerkom, *B.A.N.*, **9**, 185, 1941.
(4) P. A. G. Scheuer and M. Ryle, *M.N.*, **113**, 3, 1953.
(5) J. E. Baldwin, *M.N.*, **115**, 684, 1955.
(6) J. E. Baldwin and P. A. G. Scheuer (in preparation)
(7) S. P. Wyatt, *Ap.J.*, **118**, 304, 1953.
(8) J. E. Baldwin, *Nature*, **174**, 320, 1954.
(9) I. S. Shklovsky, *Ast. Zh.*, **29**, 418, 1952.
(10) J. R. Shakeshaft, *Phil. Mag.*, **45**, 1136, 1954.
(11) W. Baade and R. Minkowski, *Ap.J.*, **119**, 215, 1954.
(12) M. Ryle and P. A. G. Scheuer, *Proc. Roy. Soc.*, A, **230**, 448, 1955.
(13) R. Adgie (in preparation).
(14) J. R. Shakeshaft, M. Ryle, J. E. Baldwin, B. Elsmore and J. H. Thomson, *Mem. R.A.S.*, **67**, Part 3, 1956.
(15) J. D. Kraus and H. C. Ko, *Nature*, **172**, 538, 1953.
(16) R. Hanbury Brown and C. Hazard, *Nature*, **172**, 997, 1953.
(17) W. Baade at a meeting of the A.A.S., reported in *Sky* and *Telescope*, **14**,, 371, 1955.
(18) G. R. Burbidge *Ap.J.*, **123**, 178, 1956.

SPECTROPHOTOMETRIC MEASUREMENTS OF EARLY-TYPE STARS

III. Further Results and Discussion for B1 Stars*

R. Wilson

Summary

Results are given for seven B1 stars. These results are strictly comparable with those given in a previous paper of these Publications (Vol. I, No. 6) for seven other stars of type B1. Most of the paper is devoted to a discussion of the results for all fourteen B1 stars.

It is shown that spectral lines due to ions other than H, He I, He II and Mg II have luminosity effects which can be represented, to quite a high accuracy, by a single parameter ϕ_L. This is adopted as a luminosity parameter and the fourteen stars discussed are therefore classified by B1, ϕ_L.

The equivalent width of Mg II 4481 is insensitive to luminosity. This may be an ionic abundance effect since Mg II has the lowest ionization potential of the ions observed; the effect may be complicated by the fact that Mg II 4481, being a member of the fundamental series, should be susceptible to Stark broadening.

Luminosity effects in the blend-free equivalent widths of the Balmer lines Hα, Hβ and Hγ are investigated. As was to be expected, there is a negative correlation between the strength of Hγ and ϕ_L. This correlation breaks down when we include stars with emission Hα, due to the presence of incipient emission in Hγ; the results indicate that this effect may also be present, to a much lesser degree, in stars with absorption Hα.

The blend-free values of the equivalent widths of the He I lines are given and their peculiar behaviour discussed.

The basic principle of the Edinburgh method of spectrophotometry is the superposition of the unsmoothed measures of the several spectrograms of one star with a consequent reduction in grain fluctuations. This can be extended by combining the measures of a number of similar stars and it has been found convenient here to form the stars into three groups according to luminosity. These three groups have been named the supergiant, giant and dwarf groups, and the mean spectrum plots representing the B1 supergiant and the B1 dwarf are reproduced in the atlas at the end of the paper. The plots involve 28 and 29 spectrograms respectively and the standard error of the grain fluctuations has been reduced to less than one per cent.

* The full text of this paper appears in *Publications of the Royal Observatory, Edinburgh*, 2, No. 1, 1956.

CONTENTS

Printed in London by Taylor & Francis, Ltd., 1956 October